Enhanced detection for antibiotic resistance genes in wastewater samples using CRISPR-enriched metagenomic method

Yuqing Mao, Joanna L. Shisler, Thanh H. Nguyen

Keywords

Antibiotic resistance genes; Metagenomic; CRISPR

Highlights

- CRISPR-NGS can detect up to 686 more ARGs than regular NGS
- CRISPR-NGS lower the detection limit by 10-fold than regular NGS
- Clinically important ARGs missed by regular NGS can be detected by CRISPR-NGS

Abstract

The spread of antibiotic resistance genes (ARGs) in the environment is a global public health concern. To date, over 5,000 genes have been identified to express resistance to antibiotics. ARGs are usually low in abundance in environmental samples, making them difficult to detect. Metagenomic sequencing and qPCR, two conventional ARG detection methods, have low sensitivity and low throughput limitations, respectively. We developed a CRISPR-Cas9-modified next-generation sequencing (NGS) method to enrich the targeted ARGs during library preparation. The false negative and false positive of this method were determined based on a mixture of bacterial isolates with known whole-genome sequences. Low values of both false negative (2/1208) and false positive (1/1208) proved the method's reliability. We compared the results obtained by this CRISPR-NGS and the conventional NGS method for five untreated wastewater samples. As compared to the ARGs detected in the same samples using the regular NGS method, the CRISPR-NGS method found up to 686 more ARGs and up to 29 more ARG families in low abundances, including the clinically important KPC beta-lactamase genes in the five wastewater samples collected from different sources. The CRISPR-NGS method increased the average read depths by 10-fold compared with the regular NGS method. The CRISPR-NGS method is promising in ARG detection in future wastewater surveillance projects. A similar workflow can also be applied to detect other targets that are low in abundance but highly multiplexed.

1 Introduction

Antibiotic resistance is an increasing threat to global public health. It is estimated that in the next decade, if without any intervention, antibiotic resistance will cause an annual loss of 3.4 trillion US dollars in gross domestic product and cause 24 million more people to live in extreme poverty (United Nations Environment Programme, 2023). In addition to clinical settings, antibiotic resistance can also spread in soil, water, agriculture production, and the food chain (Samreen et al., 2021; Skandalis et al., 2021). Studies show that the environment plays a more important role than previously expected in antibiotic resistance transmission and evolution (Bengtsson-Palme et al., 2018; Larsson & Flach, 2021). Therefore, environmental surveillance is an essential step in antibiotic resistance stewardship (Razzaque, 2021).

Approaches for the environmental surveillance of antibiotic resistance can be culture-based or culture-independent. Culture-based methods for determining the minimal inhibitory concentration (MIC), such as broth dilution and disk diffusion, have the advantage of obtaining the real antibiotic resistance phenotypes (Kowalska-Krochmal & Dudek-Wicher, 2021). However, these types of conventional culture-based methods are labor-intensive and time-consuming (Burnham et al., 2017). Because only culturable bacteria can be tested by these methods, the MIC of antibiotics for the vast number of nonculturable bacteria in environmental samples will remain unknown with only culture-based methods (McLain et al., 2016).

Culture-independent methods are usually less labor-intensive than culture-based methods (Anjum et al., 2017; Liguori et al., 2022). Genes that encode antibiotic resistance (ARGs) are included in bacterial genomes. By extracting and analyzing the DNA extracted from environmental samples, the antibiotic resistance profiles of the entire sample can be obtained. There are two challenges to detect ARG in environmental DNA samples. First, the concentrations of many ARGs can be very low. ARGs account for less than 1% of the DNA in bacterial genomes (Debroas & Siguret, 2019). Second, there is a wide range of ARG targets. To date, more than 5,000 reference sequences of ARGs are recorded in The Comprehensive Antibiotic Resistance Database (CARD) (Alcock et al., 2023). Metagenomics and quantitative polymerase chain reaction (qPCR) are two culture-independent methods that are widely used to detect antibiotic resistance genes (ARG) (Liguori et al., 2022). These two methods have their pros and cons in ARG detection. qPCR has a high sensitivity, so it can capture the presence of ARGs that are in extremely low abundance in a sample (Bengtsson-Palme et al., 2023). However, qPCR requires the design of specific primers for each ARG target. When the number of targets increases, though multiplexed qPCR or microfluidic qPCR can be applied, the interaction among different primers and probes may become complex and make the results inaccurate (Shen, 2023). In addition, qPCR only returns fluorescent signals when short DNA sequence of 70-300 bp are detected (Ginzinger, 2002). These results can neither indicate the mutations in ARGs nor tell the existence of the remaining sequences outside of this range. Such disadvantages make qPCR easy to generate false positive results. In contrast, metagenomics does not require primer design for specific targets, so that it can detect a wide range of ARGs (de Abreu et al., 2020). However, because of the low percentage of ARGs in environmental DNA samples, most of the reads obtained in metagenomics are not relevant to ARG detection. The sensitivity of metagenomics is lower than qPCR, making the detection of low-abundance ARGs difficult by metagenomics (Liguori et al., 2022).

To balance the throughput and the sensitivity in ARG detection, different methods have been developed. Hybridization capture is one of the commercialized multiplexed target-enrichment method for next-generation sequencing (NGS). This method adds an additional

step after indexed library preparation using streptavidin magnetic beads and specifically designed biotinylated DNA probes to capture DNA fragments with matched sequences (Guitor et al., 2020). Thousands to millions of probes can be used to target various DNA sequences (Samorodnitsky et al., 2015). However, this method is expensive because of the cost of streptavidin magnetic beads and biotinylated DNA probe synthesis (Sundararaman et al., 2023). Hybridization capture often requires a high DNA sample input of microgram level which is difficult for environmental samples to achieve (Duncavage et al., 2011; Flint et al., 2023). Recently, CRISPR-Cas9-based target enrichment NGS methods were developed for clinical diagnosis (Gilpatrick et al., 2020; Quan et al., 2019). Such methods use specifically designed guide RNA (gRNA) to guide Cas9 nuclease specifically cleave the target genes to generate 5' phosphate group for adapter ligation. The libraries generated by this method can be applied to both the Illumina NGS platform and the Nanopore 3rd-generation sequencing platform (Gilpatrick et al., 2020; Quan et al., 2019). For qPCR, multiplexed primers and probes may have complex and unexpected interactions (Xie et al., 2022). In contrast, the interaction of Cas9-gRNA complexes to DNA samples is simpler, making the Cas9 cleavage reactions possible to be highly multiplexed (McCarty et al., 2020). Not like biotinylated DNA probes in hybridization capture which do not include conserved regions (Hayden et al., 2022), the gRNA used for Cas9 cleavage contains conserved regions (Liang et al., 2015), making the synthesis of multiplexed assays of CRISPR-Cas9 enriched methods more flexible and economical.

Next-generation sequencing (NGS) is increasingly accessible to laboratories because of its reduced cost (McCombie & McPherson, 2019). In this study, we developed a novel CRISPR-NGS method for ARG detection in untreated wastewater samples. This CRISPR-NGS method includes six major steps (**Figure 1**). Step 1 is to prepare multiplexed guide RNA (gRNA) targeting different sites in ARGs. The specific cleavage of Cas9 requires the specific binding of gRNA to the DNA sample. Our method used a pool of highly multiplexed gRNA (i.e., 6010 different crRNA sequences), purchasing these gRNA individually will be extremely costly. We therefore chose to synthesize tracrRNA and multiplexed crRNA from DNA templates to reduce the cost (Liang et al., 2015). Step 2 is to prepare the DNA samples for sequencing. DNA was extracted from environmental samples and blocked by alkaline phosphatase. In NGS library preparation, adapter ligation is a necessary step which requires DNA fragments have 5' phosphate. By removing the 5' phosphate using alkaline phosphatase, adapters could not be ligated to the DNA fragments. A 250-bp DNA sequence obtained from *Pseudogulbenkiania* sp. strain NH8B which should be absent in most DNA samples was synthesized with 5' phosphate and spiked in the alkaline phosphatase-treated DNA sample as the positive control for sequencing (Zhang & Ishii, 2018). Step 3 is to specifically cleave the DNA samples at the sites in the ARGs using Cas9 nuclease to generate new 5' phosphate and A-tail for adapter ligation. Step 4 is to ligate the adapters to the specifically cleaved ARG fragments. Step 5 is to amplify the adapter-ligated fragments by PCR using corresponding primers to enrich the ARG fragments in the DNA sample. Step 6 is to check the fragment lengths and the DNA concentrations of the library and load the libraries for sequencing. As described here, the CRISPR-NGS method modifies the DNA fragmentation step in NGS library preparation to increase the sensitivity of NGS in ARG detection can be largely increased. Hundreds more ARGs in low abundances in sewage samples can be uncovered. With the same total number of reads, an average of 10-fold increase in ARG read depths can be achieved in Illumina sequencing.

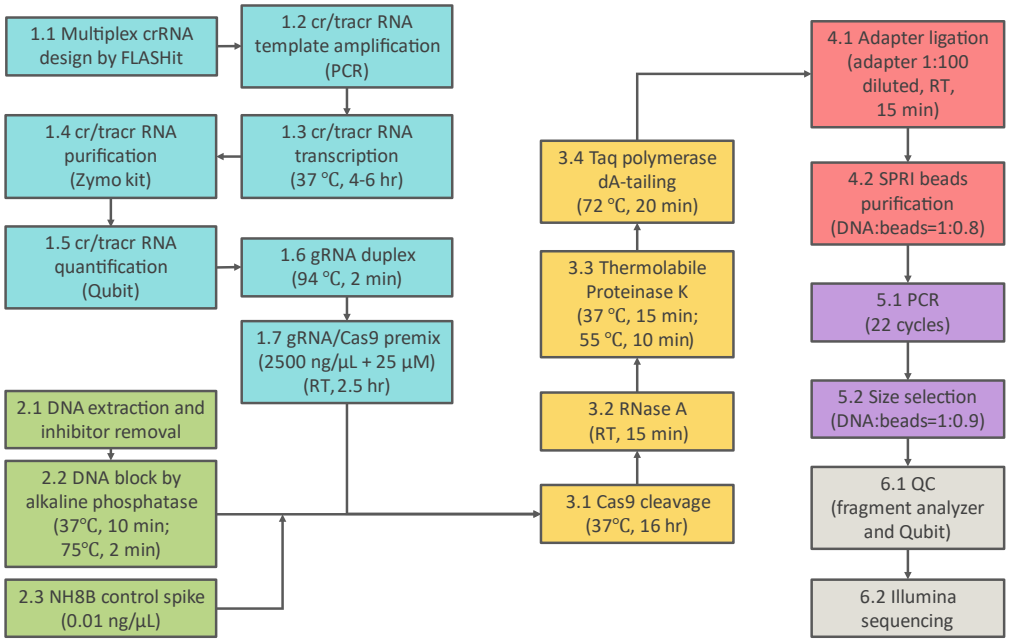

Figure 1. The workflow of the CRISPR-NGS method.

2 Methods

2.1 Environmental sample collection, pre-treatment, and DNA extraction

The environmental samples used in this study included wastewater influents from wastewater treatment plants in Chicago and raw sewage composite samples directly collected from a manhole. The wastewater treatment plant influents were taken as 24-hour composites on weekdays. The manhole sewage sample was collected by an autosampler programmed to collect four-day composite sewage samples at 4-hour intervals as described in a previous study (Oh et al., 2022). After collection, the water samples were sealed in sterile sample bags and kept in coolers filled with ice packs during transportation and stored at 4 ℃ for fewer than three days before filtering. Detailed information of the samples is listed in **Table 1**. Wastewater samples were filtered through 47 mm mixed cellulose esters (MCE) membrane filters. After filtering, the filters were cut into quarters and stored at -80 ℃ before DNA extraction. DNA was extracted from each filter using the FastDNA SPIN Kit for Soil (MP Biomedicals), using the protocol provided by the manufacturer. The centrifugation step after bead beating was increased to 15 minutes to ensure a thorough separation of liquid and solid phases. The DNA extracts were eluted by 100 µL DES solution provided in the kit. The DNA samples were purified by OneStep PCR Inhibitor Removal Kit (Zymo Research). The DNA yield was determined by Qubit™ 1X dsDNA High Sensitivity (HS) Assay Kit (Invitrogen). Then, the DNA samples were stored at -20 ℃ before library preparation.

Table 1. All sewage samples used in this study

Sample	Time	Location	Description	Volume (mL)
Calumet WWTP 1	8/16/2023	Chicago	Wastewater influent	50
Calumet WWTP 2	12/19/2023	Chicago	Wastewater influent	50
Lemont WWTP 1	8/16/2023	Chicago	Wastewater influent	75
Lemont WWTP 2	12/19/2023	Chicago	Wastewater influent	50
Campus Town manhole	11/30/2021	Central Illinois	Raw sewage from a manhole	20

2.2 Mock community DNA sample preparation

We estimated the false positive and false negative detection of ARGs using the mixture of DNA extracts at equal concentrations from 10 bacteria isolates (3 *E. coli* from wastewater, 3 *E. coli* from a clinical source, 1 Shigella from a clinical source, 2 *Salmonella* from a clinical source, and one *Pseudomonas aeruginosa* PA14 (see **Table S1**) (Grace et al., 2022). This mock community was subjected to both whole genome sequencing and CRISPR-NSG to determine the false positive and false negative rates of the CRISPR-NGS method. ZymoBIOMICS DNA Miniprep Kit (Zymo Research) was used for DNA extraction. Before DNA extraction, each bacterial strain was streaked for isolation onto culture plates from their frozen stocks and incubated at 37 ℃ for 24 hours. Then, single colonies were picked from the plates and spiked into 900 μL of LB broth and incubated at 37 ℃ overnight for bacteria enrichment. The overnight cultures were centrifuged at 600x g for 5 minutes to separate bacteria cells from the liquid medium. The pellet was re-suspended by 750 μL of ZymoBIOMICS Lysis Solution and directly transferred to ZR BashingBead™ Lysis Tubes. Further DNA isolation occurred by following the manufacturer's protocol. The DNA was eluted in 100 μL of ZymoBIOMICS™ DNase/RNase Free Water. The DNA yield was determined by Qubit™ 1X dsDNA High Sensitivity (HS) Assay Kit (Invitrogen) immediately after extraction. After mixing the DNA extracts of the 10 isolates, the mock community was stored at -20 ℃ before library preparation.

Table S1. Mock community composition

Isolate	Species/Genus	DNA conc (ng/uL)
Environmental E. coli 1	*E. coli*	>60
Environmental E. coli 2	*E. coli*	>60
Environmental E. coli 3	*E. coli*	>60
Clinical E. coli 1	*E. coli*	>60
Clinical E. coli 2	*E. coli*	>60
Clinical E. coli 3	*E. coli*	>60
Clinical Shigella 1	*Shigella* spp.	57
Clinical Salmonella 1	*Salmonella* spp.	54
Clinical Salmonella 2	*Salmonella* spp.	48.5
PA14	*Pseudomonas aeruginosa*	>60

2.3 Whole-genome sequencing and genome annotation for the bacteria isolates in the mock community

GenFind V3 (Beckman Coulter) was used to extract high molecular weight DNA from each isolate listed on Table S1. These DNA extracts were subjected to PacBio HiFi whole-genome sequencing. Similar to above, a portion of frozen bacterial stock was streaked for isolation using the same method as mentioned in section 2.2. After the plates were incubated overnight at 37 ℃, an additional enrichment step was performed by transferring 100 µL of the overnight culture into 900 µL fresh LB broth and incubating at 37 ℃ for 2-3 hours. Each DNA sample was eluted by 200 µL 10 mM Tris-HCl (pH=8). The concentrations and the lengths of the DNA extracts were determined by Qubit™ 1X dsDNA High Sensitivity (HS) Assay Kit (Invitrogen) and 1% agarose gel electrophoresis. The DNA extracts were stored at -80 ℃ before further analysis.

Before PacBio HiFi library preparation, all DNA extracts were sheared by Megarupter 3 (Diagenode) at speed 40 to obtain mean fragment lengths at 10 kbp for each sample. After shearing, the sheared DNA samples were purified by OneStep PCR Inhibitor Removal Kit (Zymo Research) followed by 1:1 beads cleaning using SMRTbell® cleanup beads (PacBio). The fragment lengths of the sheared DNA samples were determined by 1% agarose gel electrophoresis. Libraries were made using the I prep kit 3.0 (PacBio). Qubit™ 1X dsDNA High Sensitivity (HS) Assay Kit (Invitrogen) was used to determine the concentrations of the sheared DNA samples and the libraries. The fragment lengths of the libraries were determined by AATI Fragment Analyzer by Roy J. Carver Biotechnology Center before sequencing. Sequencing was also conducted by Roy J. Carver Biotechnology Center on SMRT REVIO cell (PacBio) with 30-hour movie time.

The HiFi reads of the bacteria isolates were assembled into whole genomes by Trycycler (Wick et al., 2021) using Canu, Flye, and Raven as the assemblers (Kolmogorov et al., 2019; Koren et al., 2017; Vaser & Šikić, 2021). Then, the fasta files containing all assembled contigs were uploaded to BV-BRC and annotated by its Genome Annotation service (Aziz et al., 2008; Olson et al., 2023). The records of the ARGs identified in these genomes from the CARD source were downloaded and labeled by their corresponding AROs using a customized Python code (Alcock et al., 2020). Due to the difference in the IDs of the ARGs used in BV-BRC Genome Annotation and CARD downloadable records, the remaining ARGs that were not automatically matched by the customized code were BLASTed manually to compare the sequences and labeled by corresponding AROs if the identity and coverage of the BLAST result were both greater than 95% (McGinnis & Madden, 2004). The AROs of the ARGs that were not used for crRNA design were excluded. This list of AROs was used as a reference to determine the true/false positive/negative in ARG detection for CRISPR-NGS method.

2.4 Design and synthesis of crRNA and tracrRNA

Multiplexed gRNA for ARG detection was designed by FLASHit (Quan et al., 2019). Four thousand six hundred and forty-one ARG nucleotide sequences for gRNA design input were downloaded from the CARD protein homolog model on September 12, 2022 (Alcock et al., 2020). The “all_offtargets.txt” file in FLASHit was modified to exclude potential off-targets in *Homo sapiens* and *Sus scrofa* genomes because the samples collected in this study were mainly human sewage. The default exclusion of off-targets in the *E. coli* BL21 genome in FLASHit was removed, because the Cas9 protein used in this study was commercially purchased. In total 6,010 different 20-nt spacer sequences were generated by FLASHit to cleave target ARGs into approximately 200-bp long fragments.

For financial reasons, multiplexed gRNA was assembled using crRNA and tracrRNA that were transcribed separately from corresponding DNA templates. The template for tracrRNA was a double-stranded synthesized DNA fragment (5’-AGGCGAATCAGATAATCGTTATGTCCAGACTGTATTAATACGACTCACTATAGGACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTT-3’) (Quan et al., 2019), and the template for multiplex crRNA was a single-stranded DNA oligo pool of the following format (The 20 Ns correspond to the 20-nt spacer sequences generated by FLASHit):
5’-TAATACGACTCACTATAGNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTATGCTGTTTTG-3’

A PCR step was performed to enrich both templates before transcription. Each PCR reaction included 25 μL of Phusion® High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs), 1.25 μL of 10 mM forward (cr/tracr: 5’-TAATACGACTCACTATAG-3’) and reverse (cr: 5’-CAAAACAGCATAGCTCTAAAAC-3’; tracr: 5’-AAAAGCACCGACTCGGTGCCAC-3’) primers respectively, 4 μL of 10 ng/μL multiplexed crRNA template or 1 ng/μL tracrRNA template, and filled up to 50 μL total volume using molecular biology grade water (Liang et al., 2015). The thermal cycle included an initial denaturation step at 98 ℃ for 10 seconds, 5 PCR cycles for crRNA template or 12 PCR cycles for tracrRNA template with 98 ℃ denaturation for 5 seconds and annealing at 55 ℃ for 15 seconds, and a final extension step at 72 ℃ for 1 minute. The PCR product was directly added to TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher) reaction as the DNA template for RNA transcription following the manufacturer’s instruction, followed by incubating at 37 ℃ for 4-6 hours until significant amount of white mist formed. After transcription, crRNA and tracrRNA were purified by RNA Clean & Concentrator-5 with DNase I Set (Zymo Research Corporation) following the total RNA clean-up protocol with DNase I treatment before RNA clean-up. The DNase I treatment was increased to 30

minutes to ensure thorough removal of residual DNA templates. The 100% ethanol was adjusted to 1.5 volume to increase short-length RNA yield. The RNA product was eluted by 15 μL of DNase/RNase-free water. The purified crRNA and tracrRNA were quantified by Qubit RNA Broad Range Assay Kit (Thermo Fisher) by diluting the RNA product 100-fold. The crRNA and tracrRNA were aliquoted and stored separately at -80 ℃ before use (Quan et al., 2019).

2.5 CRISPR-NGS library preparation

Instead of using non-specific enzymes for DNA fragmentation in most NGS library preparation workflows, in the CRISPR-NGS library preparation workflow, we used Cas9 nuclease to specifically cleave targeted sites for DNA fragmentation. The detailed protocol of CRISPR-NGS library preparation can be found at: https://dx.doi.org/10.17504/protocols.io.8epv5xdnjg1b/v1. Right before CRISPR-NGS library preparation, crRNA and tracrRNA were mixed equimolarly with the final concentration adjusted by Duplex Buffer (Integrated DNA Technology) to 2,500 ng/μL. The mixture was heated up to 94 ℃ for 2 minutes and slowly cooled down to room temperature for gRNA assembly (Miura et al., 2019). Then the duplexed gRNA was mixed with pre-diluted 25 μM Alt-R™ S.p. HiFi Cas9 Nuclease V3 (Integrated DNA Technologies) in a 10:3 volume ratio and incubated at room temperature for 2.5 hours to form Cas9-gRNA complex.

During Cas9-gRNA complex formation, rAPid Alkaline Phosphatase (Roche) was used to remove the 5' phosphate groups of the DNA samples. A 10 μL phosphatase reaction included 100-300 ng of DNA samples, 1 μL of rAPid Alkaline Phosphatase Buffer (10x concentrated), and 0.5 μL of rAPid Alkaline Phosphatase (1 U/μL). The mixture was incubated at 37 ℃ for 10 minutes for 5' phosphate group removal followed by 75 ℃ for 2 minutes for phosphatase inactivation.

The Cas9 cleavage reaction included 3 μL of NEBuffer™ r3.1 (New England Biolabs), 10.4 μL of the Cas9-gRNA complex, 1 μL of 0.01 ng/μL 5' phosphorylated synthesized external 250-bp DNA spike with no cleavage site in its sequence (5'-ACCCATACAAGGAACCCGGCCAGCACTACGCTCACTACGGCCGGTGGTACGGTGGGCACTCCGGTGAAATGCACGTGCTTGGCATGCCGTCAGGCCGTGAAGTCAAGCGCACCCCGGTGTTCAACATGGACAGCAACAAGATGACCATCCACATCGCCTCGCCGGCGCCGGCATACAGTCTGGGGGGAATTCAAGATGGAGAAGGGCGACGAGGTAATGGCGATCCTGACCTCGACAAGTGGAAGACCTG-3'), 6-8 μL of phosphatase treated DNA sample, and the volume was filled up to 30 μL by molecular biology grade water. The mixture was pipette mixed and incubated at 37 ℃ for 16 hours for thorough cleavage.

After the 16-hr incubation, 2.5 μL of RNase A (10 mg/mL) was directly added to the mixture and incubated at room temperature for 15 minutes. Then, 2.5 μL of Thermolabile Proteinase K (New England Biolabs) was added to each sample, and samples were incubated at 37 ℃ for 15 minutes for protein digestion, followed by 55 ℃ for 10 minutes for Thermolabile Proteinase K inactivation. The DNA fragments were dA-tailed by Taq polymerase. A master mix was premade by mixing 2 μL of dATP (100 mM), 5 μL of Taq DNA polymerase (5,000 U/mL), 80 μL of Standard Taq Reaction Buffer (New England Biolabs), and filled the volume up to 100 μL by molecular biology grade water. Five microliters of the dA-tailing master mix were added to each reaction and incubated at 72 ℃ for 20 minutes. After dA-tailing, the cleaved DNA samples were ready for adapter ligation.

We used xGen™ UDI-UMI Adapters (Integrated DNA Technologies) in the adapter ligation step to reduce the bias generated by the PCR step afterward. The adapter ligation reaction included 35 μL of the dA-tailed DNA sample, 2 μL of 100-fold diluted adapter, 30 μL of NEBNext® Ultra™ II Ligation Master Mix (New England Biolabs), and 1 μL of NEBNext® Ligation Enhancer. The reaction was incubated at room temperature for 15 minutes. After adapter ligation, 57 μL of AMPure XP Reagent (Beckman Coulter) was added to the reaction for sample clean up. The adapter-ligated DNA samples were eluted by 15 μL 0.1x TE buffer for PCR amplification.

The 50-μL PCR reaction included 15 μL of adapter-ligated DNA sample, 5 μL of xGen™ Library Amplification Primer Mix (Integrated DNA Technologies), 25 μL of NEBNext® Ultra™ II Q5® Master Mix (New England BIolabs), and 5 μL of molecular biology grade water. The thermal cycle of PCR included an initial denaturation step at 98 ℃ for 30 seconds, 22 cycles of PCR step (98 ℃ for 10 seconds and 65 ℃ for 75 seconds), and a final extension step at 65 ℃ for 5 minutes. The PCR product was cleaned by adding 45 μL of AMPure XP Reagent (Beckman Coulter) into the 50 μL reaction. The cleaned PCR product was eluted by 30 μL 0.1x TE buffer as the CRISPR-NGS library. The libraries were stored at -20 ℃ before sequencing.

2.6 Regular NGS library preparation

For a regular NGS library preparation protocol, there are four major steps before sequencing: DNA fragmentation, adapter ligation, PCR (optional), and quality control. Instead of using a random enzyme digestion step used in most NGS library preparation kits, the CRISPR-NGS method applies the specific cleavage feature of CRISPR-Cas9 to selectively cleave the target ARGs into fragments with lengths of hundreds of base pairs as the DNA fragmentation step. The regular NGS libraries were prepared using NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina (New England Biolabs). The same amount of DNA samples as the input of the CRISPR-NGS libraries were used to make the regular NGS libraries. One microliter of 0.01 ng/μL 5' phosphorylated synthesized external 250-bp DNA spike as mentioned in the above section was added (Zhang & Ishii, 2018). The library preparation steps followed manufacturer's instructions with minor adjustments. In the fragmentation step, the samples were incubated at 37 ℃ for 30 minutes. xGen™ UDI-UMI Adapters (Integrated DNA Technologies) was used for adapter ligation, so that the USER® Enzyme incubation step was skipped. The adapter dilution ratios were adjusted by the DNA input for each library (**Table S2**). Five cycles of PCR were used in the PCR enrichment step. The final library was eluted in 30 μL of 0.1x TE buffer and stored at -20 ℃ before sequencing.

Table S2. Information for library preparation

Method	Library	Adapter dilution ratio	DNA input (ng)	NH8B spike amount (ng)
CRISPR-NGS	Mock community	1:100	123.8	8.66E-03
	Calumet 8/16/2023	1:100	115.2	8.06E-03
	Lemont 8/16/2023	1:100	121.2	6.86E-03
	Calumet 12/19/2023	1:100	110.4	6.86E-03
	Lemont 12/19/2023	1:100	136.8	6.86E-03
	Campus Town 11/30/2021	1:100	65.8	8.06E-03
Regular NGS	Mock community	1:1	122.9	7.60E-03
	Calumet 8/16/2023	1:1	108.0	1.11E-01
	Lemont 8/16/2023	1:1	161.2	7.60E-03
	Calumet 12/19/2023	1:1	166.4	7.60E-03
	Lemont 12/19/2023	1:1	200.6	7.60E-03
	Campus Town 11/30/2021	1:10	100.1	7.60E-03

2.7 Library quality control and next-generation sequencing

The concentrations of the libraries were determined by Qubit™ 1X dsDNA High Sensitivity (HS) Assay Kit (Invitrogen) right after library preparation. After concentration check, the libraries were sent to Roy J. Carver Biotechnology Center for following quality check steps and sequencing. DNA fragment lengths were determined by AATI Fragment Analyzer. An additional beads purification was conducted before sequencing to remove excess adapter dimers according to the fragment analysis results. A small portion of the libraries were pooled and pre-sequenced on Illumina MiSeq V3 to determine the optimal pooling ratio. Then, all libraries were pooled according to the MiSeq read depths and sequenced on Illumina NovaSeq X Plus 10B paired-reads.

2.8 Sequencing read mapping

After paired-end sequencing, the UMI indexes were extracted into a separated fastq file. For each library, three fastq files were generated to store read 1, read 2, and the UMI indexes. All fastq files were sorted by the command “cat file.fastq | paste - - - - | sort -k1,1 -t " " | tr "\t" "\n" > file_sorted.fastq” before processing. UMI-Tools was used to add UMI indexes from the separated UMI index file to corresponding identifiers in read 1 and read 2 files using the command “umi_tools extract --extract-method=string --bc-pattern=NNNNNNNNN -I [UMI_index_fastq_file] -S [UMI_index_fastq_file_out_nothing_left] --read2-in=[read_fastq_file_to_add_UMI] --read2-out=[output_read_fastq_file_with_UMI_added]” (Smith et al., 2017). After adding UMI indexes, the sequencing reads in read 1 and read 2 files were screened by PriceSeqFilter using the criteria that 85% of nucleotides in a read must be high quality, minimum allowed probability of a nucleotide being correct is 0.98, and 90% of nucleotides in a read that must be called (Ruby et al., 2013). The output fastq files from PriceSeqFilter were used for read mapping in Bowtie2 with very sensitive end-to-end mapping and reporting all alignments (Langmead & Salzberg, 2012). A fasta file including the sequences of all ARGs for crRNA design and the sequence of the 250-bp external DNA spike was used as the reference. The output sam files were converted to bam files, sorted, and indexed by Samtools (H. Li et al., 2009). The bam files were deduplicated using UMI-Tools according to the UMI indexes added to the read identifiers earlier (Smith et al., 2017). The coverage and average read depths of the mapping for all ARGs and the external DNA spike were exported by the coverage function in Samtools with “-ff 0” setting to remove the restriction on secondary alignment (H. Li et al., 2009). Any ARGs with coverages higher than 70% were considered present in the sample. The threshold of 70% was selected by comparing the length of cleaved part of the ARGs to the full ARG lengths, to include as many ARG targets as possible. Among the 4641 ARGs used for gRNA design, 273 ARGs had designed cleaved part lower than 70% of their full lengths, 1037 were between 70% and 80%, 1186 were between 80% and 90%, and 2145 were between 90% and 100%.

After the above steps, a validation step was conducted to check the reliability of read mapping. In all ARG families that were detected by sewage CRISPR-NGS libraries, one ARG from each family was randomly selected and viewed for these CRISPR libraries in Integrative Genomics Viewer (IGV) (Robinson et al., 2011). Since DNA in CRISPR-NGS libraries were fragmented by cleaving at designed sites and DNA in regular NGS libraries were fragmented randomly, sharp boundaries of reads were expected at cleavage sites for CRISPR-NGS libraries while the read mapping histograms were smooth for regular NGS libraries (**Figure S1**). Read mapping was checked by comparing the cleavage sites shown in IGV to the bed file generated after gRNA design indicating all possible cleavage sites for all ARGs. The ARGs with reads covered from the start to the end of the cleavage sites were

labeled fully covered. For example, the length of an ARG is 739 bp, with its first cleavage site between 85 and 108 bp and the last cleavage site between 706 and 729 bp. Even if there are no reads mapped upstream of 85 bp or downstream of 729 bp, this ARG is considered fully covered in a CRISPR-NGS library because the sequence between these two sites is fully covered by the reads starting and ending at the other designed cleavage sites (**Figure S1**).

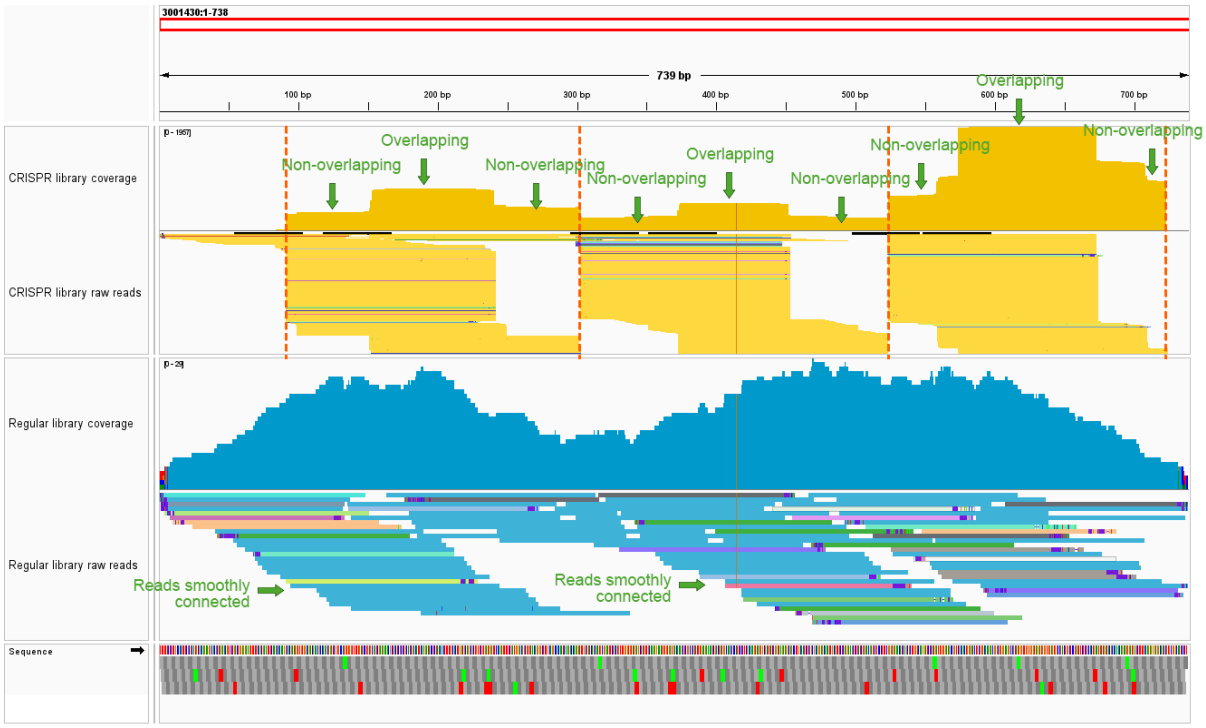

Figure S1. The difference in read mapping between the CRISPR-NGS library (top) and the regular NGS library (bottom) using the example of OXA-36 (ARO: 3001430). The number axis at the top of the figure shows the scale of OXA-36 in base pairs (bp). The DNA fragments were cleaved at the four gRNA binding sites (85-108 bp, 296-319 bp, 517-540 bp, and 706-729 bp, labeled by orange dash lines) for the CRISPR-NGS library.

2.9 ARG clustering for quantification

Among the 4641 ARGs used for gRNA design, some have highly similar sequences with only a few mutations. One read may be mapped to multiple ARGs that share highly similar sequences. To avoid the bias in ARG quantification due to the overlaps in read mapping, we divided the 4641 ARGs into 1208 clusters based on their average nucleotide identities (ANI). The pairwise ANIs for all ARGs were calculated by FastANI using fragment length of 100 bp and minimum fraction of a sequence that must be shared to be 100% (Jain et al., 2018). The FastANI output was screened by a customized python code to keep only the ANIs above 90%. A network was built for the screened ANIs in Cytoscape (Shannon et al., 2003). The ARGs were plotted as nodes. Two nodes were connected if the ANI between them was greater than 90%. The connected nodes were defined to be in the same cluster. If a node did not connect to any other nodes, it was in an individual cluster. There were 228 clusters containing more than one ARGs, and 980 clusters containing only one single ARG. The read depth of a cluster was calculated by the average read depths of all ARGs in this cluster.

2.10 Sanger sequencing for ARG presence validation

Sanger sequencing was conducted to validate the presence of the ARGs detected in the DNA samples. The ARGs targeted in this validation step were randomly selected from the ARG mapping results. The sequences of the selected ARGs were downloaded from CARD (Alcock et al., 2023). Then, these sequences were used as the input for Primer BLAST on NCBI to design primers for long-range PCR (Ye et al., 2012). The minimum PCR product size was modified to be less than 100 bp smaller than the size of the input ARG to ensure covering the largest possible segment of the gene. The database used in Primer BLAST was nr, and the organism was set as bacteria (taxid: 2). The long-range primers are listed in **Table S3**.

Table S3. The PCR primers designed for Sanger sequencing validation

#	Gene	Gene Length	Amplicon Length	Forward	Reverse
1	MCR-4.1	1626	1401	CACTGCGTTGTTTTATGTTGCC	GCCAGCATTGGTACGCTAGT
2	TEM-1	861	827	TTCGTGTCGCCCTTATTCCC	TGAGGCACCTATCTCAGCGA
3	aac(3)-IId	861	810	TGCATACGCGGAAGGCAATA	TCGGAGTGGTTCCGAAATGC
4	tet(G)	1176	1101	ATCATTGCCCTGCTGATCGT	GCCGACAATCCAAACCCAAC
5	qnrB5	681	605	ACTCTGGCATTAGTTGGCGA	AACGATGCCTGGTAGCTGTC
6	GES-26	864	829	GCGCTTCATTCACGCACTATT	AACTTGACCGACAGAGGCAA
7	aph(3")-Ib	804	780	GGTGAATCGCATTCTGACTGGTT	TCAACCCCAAGTCAGAGGGT
8	MCR-5.1	1644	1609	CGGTTGTCTGCATTTATCACTTTCT	GAAGGTCCAACTCTGGCGT
9	sul1	840	833	GTGACGGTGTTCGGCATTCT	GCATGATCTAACCCTCGGTCT
10	CRP	633	607	GACCCGACTCTCGAATGGTT	AACGAGTGCCGTAAACGACG
11	sul2	816	800	TCGCTCATCATTTTCGGCATC	ATTCTTGCGGTTTCTTTCAGCG
12	aph(3')-Ia	816	806	ATGAGCCATATTCAACGGGAAAC	TCATCGAGCATCAAATGAAACTGC
13	dfrA1	474	454	TGGTAGCTATATCGAAGAATGGAGT	CCCTTTTGCCAGATTTGGTAACTAT
14	MCR-5.1_latter_part	1644	778	GACGGATACGGCTACATCCC	TCTGCATGTTGCCAGAAGGT
15	MCR-4.1_latter_part	1626	504	TGGCAGTAGCACCAAGTCAA	GGCTCTGATAGACGGTGGTT

For long-range PCR, we added 12.5 μL of Phusion® High-Fidelity PCR Master Mix (New England Biolabs) or NEBNext® Ultra™ II Q5® Master Mix (New England BIolabs), 1.25 μL of 10 μM forward and reverse primers respectively, 1 μL of the DNA sample, and filled the volume up to 25 μL by nuclease-free water for each reaction. The PCR thermal

cycle includes an initial denaturation step of 98 ℃ for 30 seconds, 34 cycles of the PCR step with 10 seconds of denaturation at 98 ℃, 30 seconds of annealing at 55 ℃, and 90 seconds of elongation at 72 ℃, and a final extension step at 72 ℃ for 5 minutes. The amplicon lengths and DNA concentrations were validated by DNA electrophoresis and Qubit™ 1X dsDNA High Sensitivity (HS) Assay Kit (Thermo Fisher) before Sanger sequencing. Sanger sequencing was conducted by UIUC Core Sequencing Facility. Each DNA sequence was aligned to the corresponding reference gene sequence in CARD by NCBI nucleotide BLAST to test their similarity (McGinnis & Madden, 2004).

2.11 qPCR for ARG quantification validation

Nine ARGs with various read depths detected by CRISPR-NGS or regular NGS methods were selected for qPCR validation. The primers for these ARGs and their corresponding amplification efficiencies are listed in **Table S4**. A 15-μL qPCR reaction included 7.5 μL of PowerUp SYBR Master Mix (Applied Biosystems), 600 nM of forward and reverse primers, 2 μL of 10-fold diluted sewage DNA extract, and 4.3 μL of molecular biology grade water. The samples were loaded in triplicates in a qPCR run. Reactions targeting the 16S rRNA gene for the same samples were also loaded in triplicates on every qPCR plate for relative abundance quantification. qPCR was performed on QuantStudio 7 Pro (Thermo Fisher) using fast cycle followed by a melt curve stage for the quality control of non-specific amplification. The qPCR thermal cycle included an initial hold stage at 95 ℃ for 20 seconds and 40 PCR cycles with 95 ℃ for 1 second and 60 ℃ for 20 seconds. After the qPCR run, the relative abundances of the ARGs were determined by ΔCq method using the following equation (Rao et al., 2013):

$$Relative\ abundance = 2^{Cq_{16S} - Cq_{ARG}}$$

In the equation, Cq_{16S} means the average Cq value of the 16S rRNA gene, and Cq_{ARG} means the average Cq value of the ARG. After obtaining the qPCR results for the nine ARGs in all sewage samples, the read depths in both CRISPR-NGS and regular NGS against the relative abundances in qPCR were plotted and fitted as linear regression curves in log10 scale in Origin Pro 2024 to evaluate the ARG quantification potential for the CRISPR-NGS method.

Table S4. Information for primers used for qPCR

Gene	Forward	Reverse	Length	Efficiency (%)	Reference
dfrA	ACGGATCCTGGCTGTTGGTTGGACGC	CGGAATTCACCTTCCGGCTCGATGTC	237	89.4	(Le et al., 2016)
sul1	CACCGTTGGCCTTCCTGTAA	TCTAACCCTCGGTCTCTGGC	180	98.4	(Le et al., 2016)
sul2	GCGCTCAAGGCAGATGGCATT	GCGTTTGATACCGGCACCCGT	293	94.3	(Le et al., 2016)
GES-26	CTTGAGAAGCTAGAGCGCGA	GCAGGAGACCATTCGACGAT	224	93	This study
MCR-5.1	CTTGCCCGTACCATTGACCT	ACAGGCTTGGTCGGCATAAA	198	101.3	This study
aph(3')-Ia	ACCGAGGCAGTTCCATAGGA	ACCTTATTTTTGACGAGGGGAAA	104	99.1	(Rochegüe et al., 2021)
KPC	CACTGTGCAGCTCATTCAAGG	CGCCGATAGAGCGCATGAA	270	100.8	(Lund et al., 2018)

OXA-48	CTTAAACGGGCGAACCAAGC	GTTCATCCTTAACCACGCCC	94	101.9	(Lund et al., 2018)
CTX-M	CGGGCRATGGCGCARAC	TGCRCCGGTSGTATTGCC	105	87.3	(Roschanski et al., 2014)
16S	CCTACGGGAGGCAGCAG	ATTACCGCGGCTGCTGG	194	97.5	(Muyzer et al., 1993)

3 Results and Discussions

3.1 Method validation

We validated that alkaline phosphatase blocking inhibited the T4 DNA ligase from connecting different DNA fragments together. After phosphatase blocking, the 5' phosphate of DNA fragments is removed, and T4 ligase could not anneal DNA fragments to form new fragments of other sizes (**Figure S2**). We used a synthetic 731-bp DNA template with 5' phosphorylation to test the blocking efficiency of alkaline phosphatase. After phosphatase blocking, the DNA was incubated with T4 ligase. As shown in **Figure S2**, while the DNA sample without blocking showed multiple bands, the blocked DNA remained its original 731 bp size, suggesting the 5' phosphate groups of blocked DNA fragments were removed and not ligated.

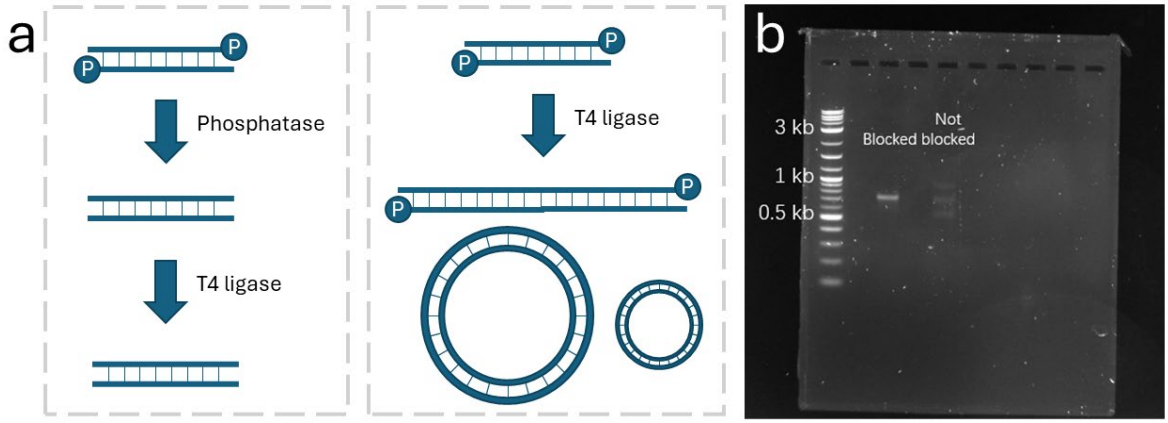

Figure S2. Validation of alkaline phosphatase blocking to remove 5' phosphate. a: The concept figure of 5' phosphate removal validation using T4 ligase. b: The gel image of blocked and unblocked samples after T4 ligase treatment.

We validated the Cas9 cleavage using an in-silico analysis that compares the breakpoints in raw sequencing reads with the intended Cas9 cleavage sites. In read mapping, paired 150-bp raw sequencing reads were aligned to the reference ARG sequence obtained from CARD. The paired reads, which started from different ends of the same DNA fragment, should have an overlapping region if the DNA fragment is smaller than 300 bp. If the reads are from different fragments, there would be no overlap. Because most DNA fragments were designed to be around 200 bp using a protocol developed previously (Quan et al., 2019) and the maximum read length was 150 bp in NGS, the entire DNA fragment would not be covered by one single read but could be covered by two paired reads from different directions. Therefore, in the read mapping plot of a cleaved 200-bp DNA fragment,

we expected to see the middle part (usually around 100 bp) overlapped by the paired 150-bp reads from different directions and the remaining left and right part covered by the remaining lengths of the paired reads. Because the DNA fragments were generated by Cas9 cleavage at a few designed sites, these DNA fragments should not overlap with each other in the read mapping of the ARG. The locations where two DNA fragments meet each other should be the same as the designed Cas9 cleavage sites. The example of overlapping regions on different DNA fragments in the same ARG was marked and shown in **Figure S1,** which shows the read mapping plot for CRISPR-NGS and regular NGS for the same gene OXA-36 (ARO: 3001430) (Alcock et al., 2023). The four cleavage sites on OXA-36 broke the read mapping histogram of CRISPR-NGS into three overlapping regions while the regular NGS read mapping histogram was smooth. We checked the read mapping of 79 ARGs from different ARG families, 78 of the raw reads were broken at the designed cleavage sites except for *rpoB* which contained a few off-target cleavage sites. This finding suggested most of the ARGs detected by CRISPR-NGS truly existed in the samples.

3.2 CRISPR-NGS method showed high accuracy in ARG detection in the mock community sample

We conducted whole-genome sequencing of 10 bacteria isolates (listed in **Table S1**) and these genomes served as a basis to determine false positive and false negatives in ARG detection by the CRISPR-NGS method. Because some ARGs share highly similar sequences, we divided the 4641 targeted ARGs into 1208 clusters based on their similarities in terms of ANI to avoid overcounting these ARGs (**Figure S3**). We first used annotation for the 10 whole-genome sequences to find 100 ARGs in 99 clusters existing in the mock community. The CRISPR-NGS method detected 1302 ARGs in 124 clusters in the same sample (**Spreadsheet S1**). **Table S5** summarizes the results. Ninety-seven clusters were found by both CRISPR-NSG and annotation of the whole genomes. These ARG clusters have identities higher than 90% with the reference genome sequence. Because the algorithm of whole-genome annotation may miss some ARGs (Maboni et al., 2022), we used BLAST to search the ARG sequences listed in the CARD database within the whole genomes of the bacterial isolates of the mock community. We found 25 additional clusters of ARGs that found by the CRISPR-NGS method in the whole-genome sequences of the mock community. Finally, we used Sanger sequencing to confirm the presence of two clusters of ARGs, one containing quinolone resistance protein (*qnr*) genes and the other containing one single ARG *aph(3')-Ia.* The amplicon of the *qnrB5* gene was aligned to the reference sequence in 99.65% identity, suggesting this gene was missed by long-read PacBio sequencing but could be detected by short-read Illumina sequencing. In total, we confirmed the presence of 125 ARGs cluster found by the CRISPR-NSG method (97 from annotation of WGS, 25 from the BLAST method and two from Sanger sequencing). These clusters are the true positive. There were two fault negatives, which were found by whole genome sequencing but not by CRISPR-NGS. These two ARG clusters that only had an identity of 86-87% with the reference genome sequences and this may be the reason causing CRISPR-NGS to miss them.

We also compared the ARG detection results for the mock community sample to the regular NGS method using the similar methods as described above. The regular NGS library detected two more ARG clusters, *dfrA14* and *aadA13*, than the CRISPR-NGS library. However, the identities of these two ARGs were 89% against the reference sequence. To conclude, CRISPR-NGS may miss some ARGs with accumulated mutations, its overall accuracy in ARG detection is high.

Table S5. Summary of positive and negative results in ARG detection using CRISPR-NGS method for the mock community sample

CRISPR ARG cluster	True	False
Positive	123 (97 from WGS, 25 from BLAST, 1 from Sanger sequencing)	1 (aph(3')-Ia)
Negative	NA	2 (identity < 90% in WGS annotation)

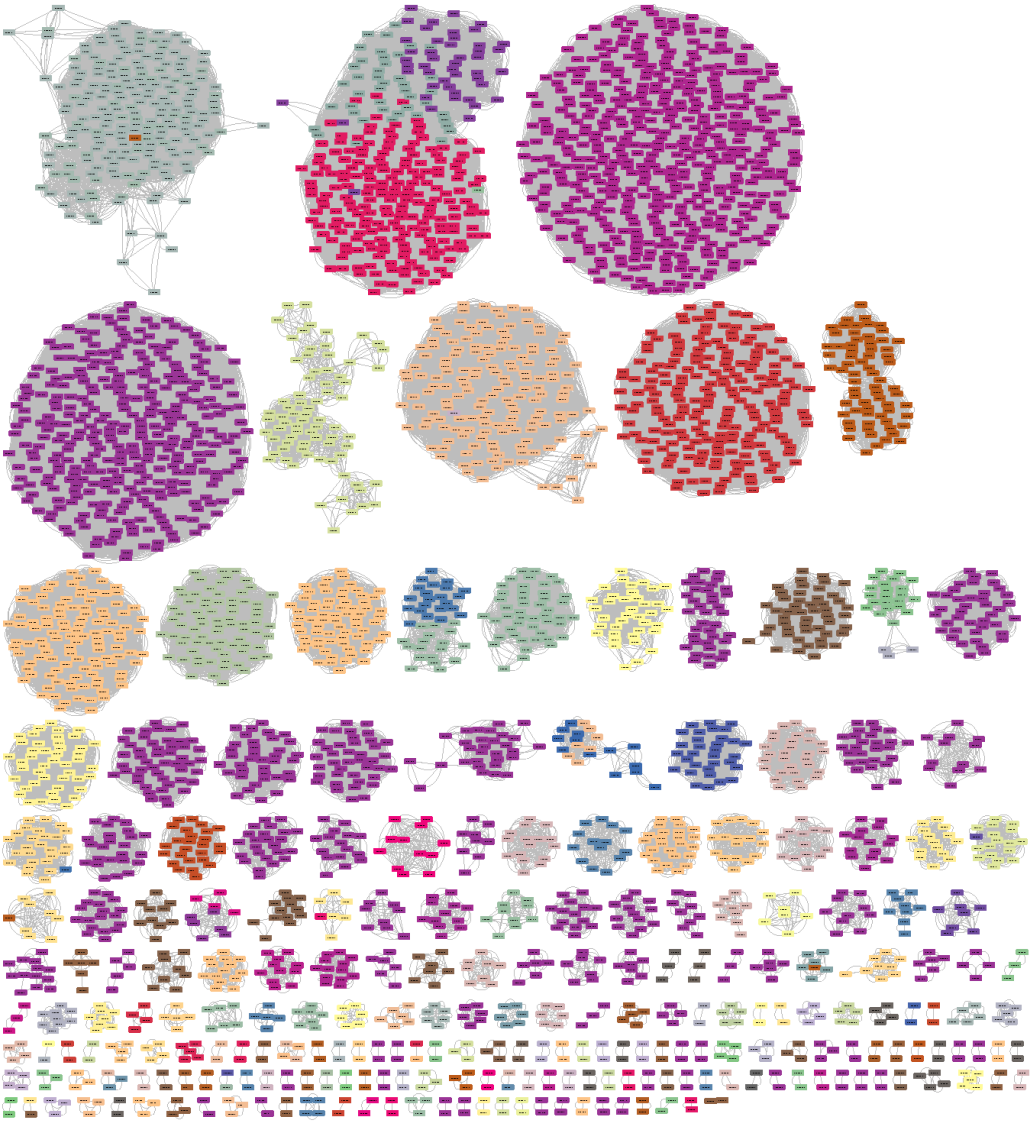

Figure S3. ARG clusters containing more than one ARGs

3.3 More ARGs were detected by the CRISPR-NGS method than regular NGS method

The improved sensitivity of the CRISPR-NGS method in ARG detection was evaluated by comparing the number of ARGs and ARG families detected in the same sewage sample by the CRISPR-NGS method and regular NGS method, respectively. **Figure 2** shows the number of ARGs and ARG families that were mapped with over 70% coverage

by only the CRISPR-NGS method, only the regular NGS method, and both methods. The five human sewage samples, including the four WWTP influent samples and the one raw sewage sample collected from a manhole, had 226-370 ARGs detected by regular NGS libraries, and these numbers increased to 428-982 for CRISPR-NGS libraries. Only between 6 to 56 ARGs were uniquely detected by regular NGS libraries, while 215-686 ARGs were uniquely detected by CRISPR-NGS libraries. These additional ARGs detected by CRISPR-NGS libraries were mostly from ARG families that were not covered by regular NGS libraries. Nineteen to 29 more ARG families were only detected by CRISPR-NGS libraries for the five sewage samples, while only 0-2 were only detected by regular NGS libraries. These ARG families detected only by CRISPR-NGS libraries include KPC and CTX-M beta-lactamase genes which are clinically important and MCR phosphoethanolamine transferase genes which encode resistance to colistin, one of the last-resort antibiotics (Han et al., 2020; McKenna, 2013).

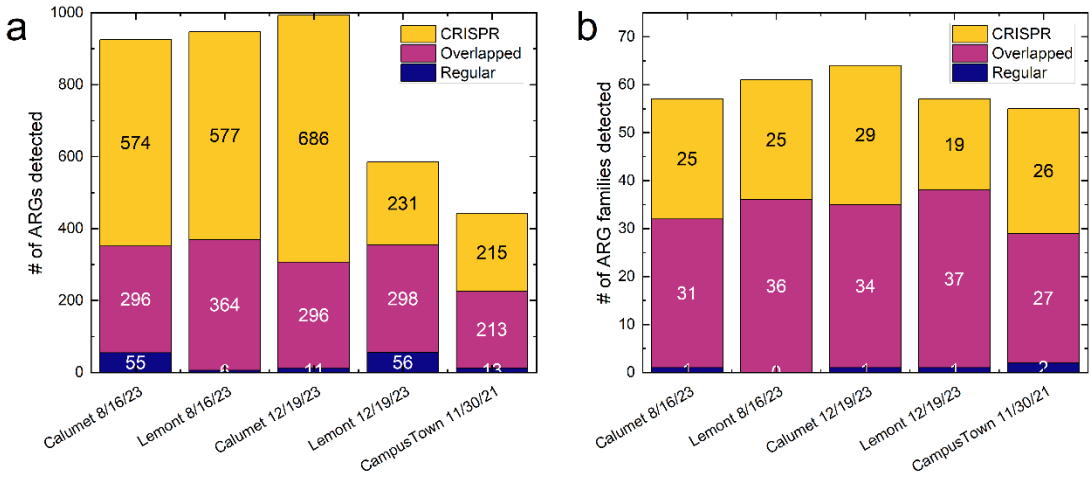

Figure 2. More ARGs (a) and ARG families (b) were detected by CRISPR-NGS method than regular NGS method for human sewage samples. The yellow bars show the number of ARGs and ARG families only detected by CRISPR-NGS method, the purple bars show the number of ARGs and ARG families detected by both CRISPR-NGS and regular NGS methods, and the dark blue bars show the number of ARGs and ARG families only detected by the regular NGS method.

3.4 CRISPR-NGS method could detect ARGs in higher read depths than regular NGS method

The ARGs detected by CRISPR-NGS libraries also had higher read depths than those detected by regular NGS libraries. The fold changes in read depths of the 159 ARGs detected in all five human sewage samples by both CRISPR and regular NGS libraries are plotted in **Figure 3**. One hundred and forty-four ARGs had significantly higher read depths in CRISPR-NGS libraries than regular NGS libraries. The fold changes in read depths of these 144 ARGs varied from 2.04 to 71.96, suggesting the CRISPR-NGS method can detect ARGs in higher confidence than regular NGS method.

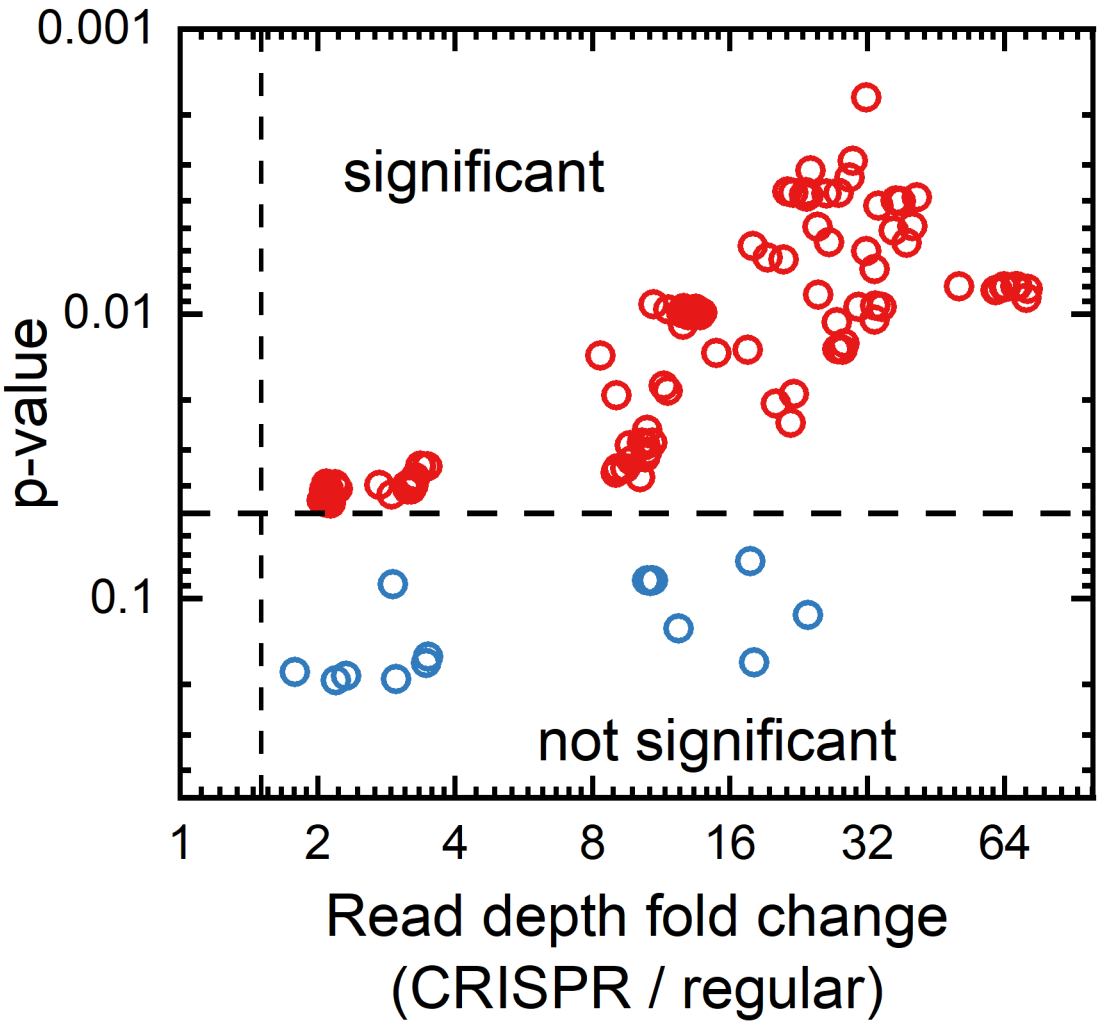

Figure 3. The read depths of 144 (red dots) in 159 ARGs detected in all human sewage samples were significantly higher in CRISPR-NGS libraries than regular NGS libraries. The read depths for the remaining 15 ARGs (blue dots) were higher but not significant. The y axis shows the values of the p-values in a reversed log-10 manner, and the x axis shows the fold changes in read depths (CRISPR-NGS versus regular NGS). The horizontal dash line means the threshold of a significant p-value equals to 0.05, and the vertical dash line means the thresholds in fold change of 1.5.

3.5 ARG presence could be validated by Sanger sequencing

We used Sanger sequencing to validate the presence of selected ARGs in sewage samples (**Table 2**). The ARG targets were amplified by PCR, and the PCR amplicons were Sanger sequenced and compared to the reference sequences in CARD. We chose Sanger sequencing for validation, because Sanger sequencing the PCR amplicons is accurate, cost-effective, and fast when there are only a few targets. The identities between the sequenced PCR amplicons and the reference sequence of the ARGs in CARD were between 89% and 100%, suggesting the amplicon sequences were highly similar to the reference sequences. Because the PCR primers designed for different ARGs had different binding regions, the

amplicons could not cover the full lengths of the ARGs. After adjusting with the amplicon lengths, the coverage of the ARGs were between 58% and 99%. The high coverages and identities of Sanger sequencing also matched the read mapping plots. For example, the two colistin resistance genes, *MCR-5.1* and *MCR-4.1*, were nearly fully mapped by CRISPR-NGS reads (**Figure S4**), though these two genes were barely detected by regular NGS.

Table 2. The Sanger sequencing identities and adjusted coverages for the ARGs detected in different CRISPR-NGS libraries

Gene	**Identity (%)**	**Coverage adjusted by amplicon lengths (%)**
aph(3'')-Ib	99-100	95-98
sul1	97-99	95-96
CRP	89-94	95
sul2	95-99	92-96
aph(3')-Ia	95-99	73-97
dfrA1	96-97	93-95
GES-26	95-100	96-97
MCR-5.1	98	77-99
MCR-4.1	98	98
TEM-1	99	96
aac(3)-IId	96-99	58-93
tet(G)	94	88
qnrB5	100	97

Note: The coverages were adjusted by dividing the aligned length by the amplicon length, rather than the coverage provided in BLAST, which divides the aligned length by the query length (i.e., the reference ARG length in CARD).

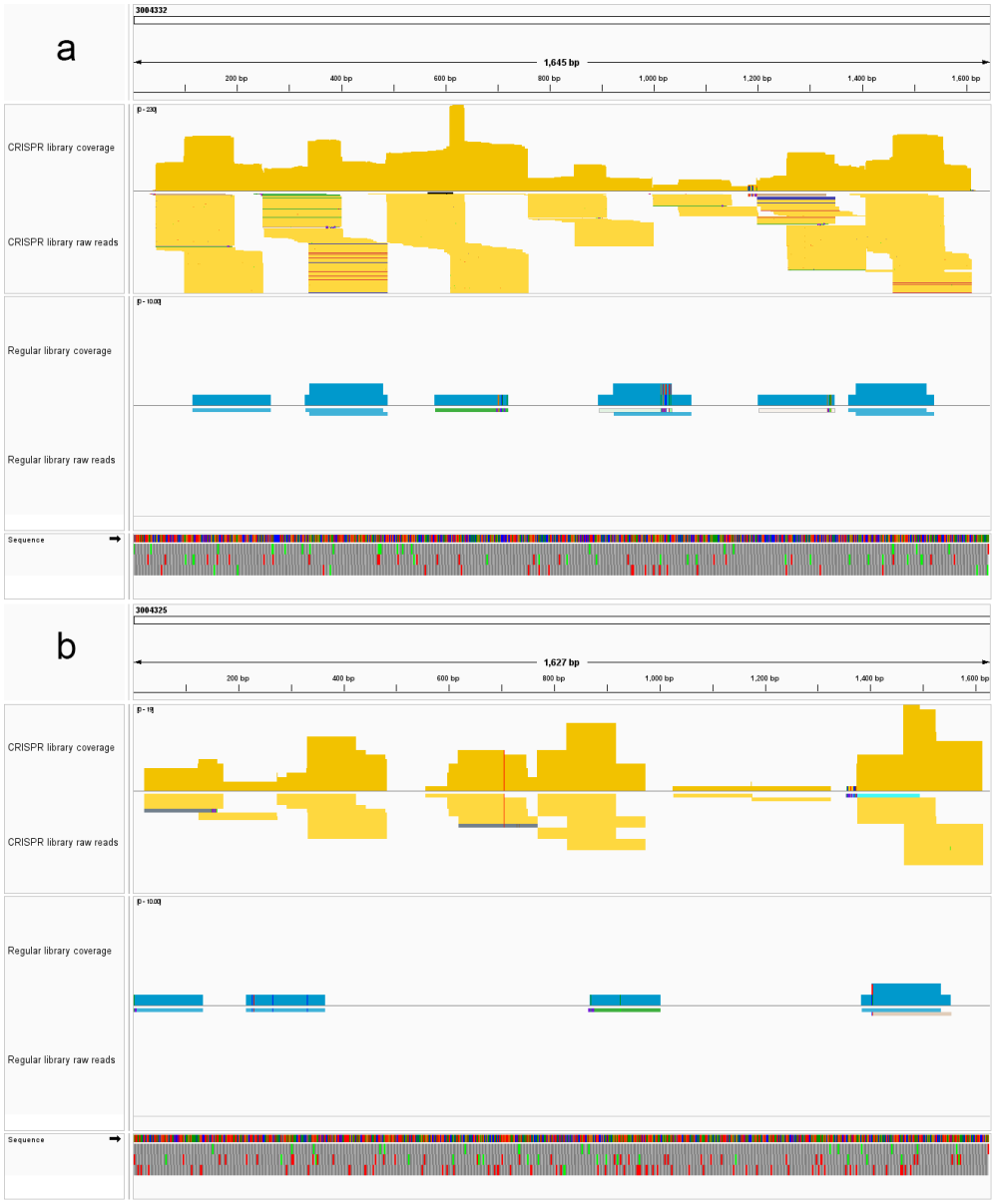

Figure S4. The read mapping of *MCR-5.1* (ARO: 3004332) (a) and *MCR-4.1* (ARO: 3004325) (b) in the CRISPR-NGS library (yellow) and the regular NGS library (blue) of Lemont (8/16/2023) sample.

3.6 CRISPR-NGS libraries had lower detection limit than regular NGS libraries, and the read depths in CRISPR-NGS libraries were proportional to the relative abundances determined by qPCR

We evaluated the quantification potential of the CRISPR-NGS method by plotting the read depths against the relative abundance detected by qPCR for nine ARGs with various read depths detected by CRISPR-NGS or regular NGS methods in the five sewage samples (**Figure 4**). The ARGs and corresponding qPCR primers are listed in **Table S4**. The read depths in both CRISPR-NGS and regular NGS libraries increased when the qPCR relative abundances increased. The read depth versus qPCR relative abundance curves for both CRISPR-NGS and regular NGS methods could be fitted linearly in $\log_{10}$ scale ($R^2 = 0.579$ and 0.848, respectively), suggesting the read depths in CRISPR-NGS method could be used

for ARG abundance comparison. In addition, the lower detection limit of the CRISPR-NGS method was down to 10^{-5} copies ARG per copy of 16S rRNA gene by, while the lower detection limit of regular NGS method was 10^{-4} copies ARG per copy of 16S rRNA gene. Thus, the CRISPR-NGS method has a higher sensitivity for CRISPR-NGS method than regular NGS method in ARG detection. A previous study reported varied relative abundances of ARGs in wastewater samples ranging from 1.4×10^{-5} to 0.09 (Srathongneam et al., 2024), suggesting the improved lower detection limit of CRISPR-NGS can promote the detection of ARGs in low abundances in wastewater samples.

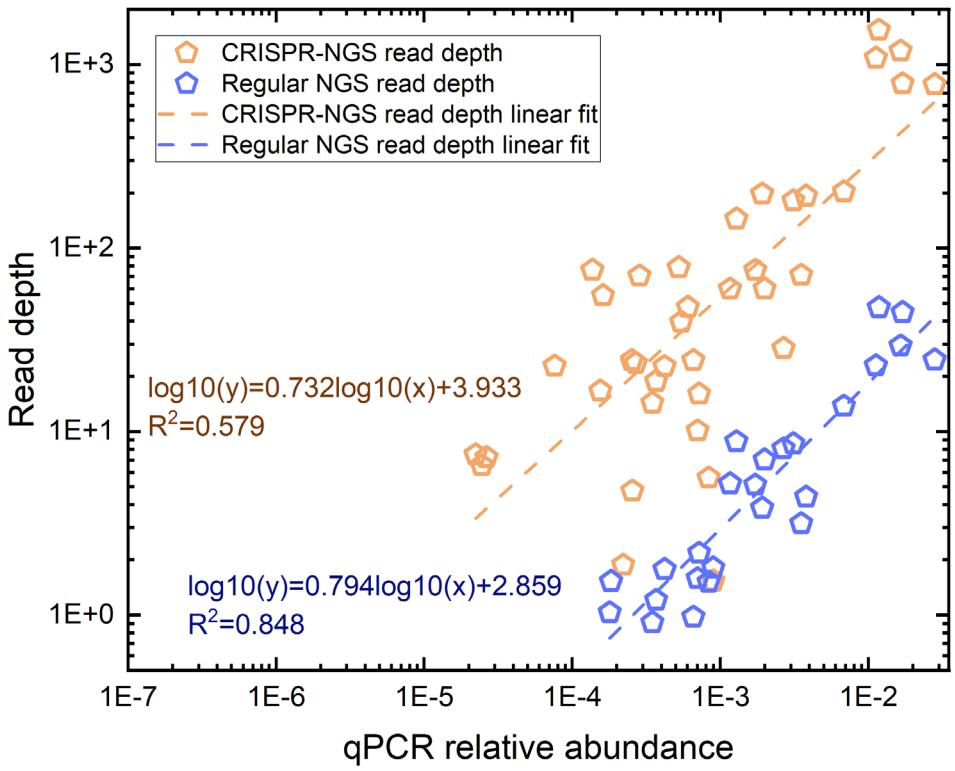

Figure 4. The regression curves of the read depths versus the qPCR relative abundances of the nine ARGs detected in different wastewater samples. The orange dots are the read depths of the ARGs detected in the CRISPR-NGS libraries. The blue dots are the read depths of the ARGs detected in the regular NGS libraries. The dash lines of corresponding colors are the linear fits of the read depths versus relative abundances for different NGS methods.

3.7 The distribution of ARGs depend on the sample sources

The most abundant clusters of ARGs (i.e., ARG clusters with read depths greater than 100 in any of the CRISPR-NGS libraries) in the five sewage samples were selected and plotted into heatmaps for comparison in **Figure 5**. These 36 ARG clusters were distributed in seven major drug classes, including aminoglycoside (9), beta-lactam (11), antiseptics (2), macrolide (5), multi-drug (3), phenicol (2), and tetracycline (4). The multi-drug class was defined to include the ARGs that are resistant to at least two drug classes in the 61 sub-terms under the CARD “antibiotic molecule” ontology term (Alcock et al., 2023). Comparing

the wastewater treatment plant sewage samples, the campus town manhole sample had lower abundances of beta-lactamase genes than the WWTP influent samples, especially for beta-lactamases MOX-9, OXA-504, and OXA-780. The abundance of the *tet(C)* gene in the campus town manhole sample was greater than any of the four WWTP influent samples, but *tet(E)* in the campus town manhole sample was lower than the four WWTP influent samples. The above findings suggested different ARG distributions in different geographical areas or different sampling scales.

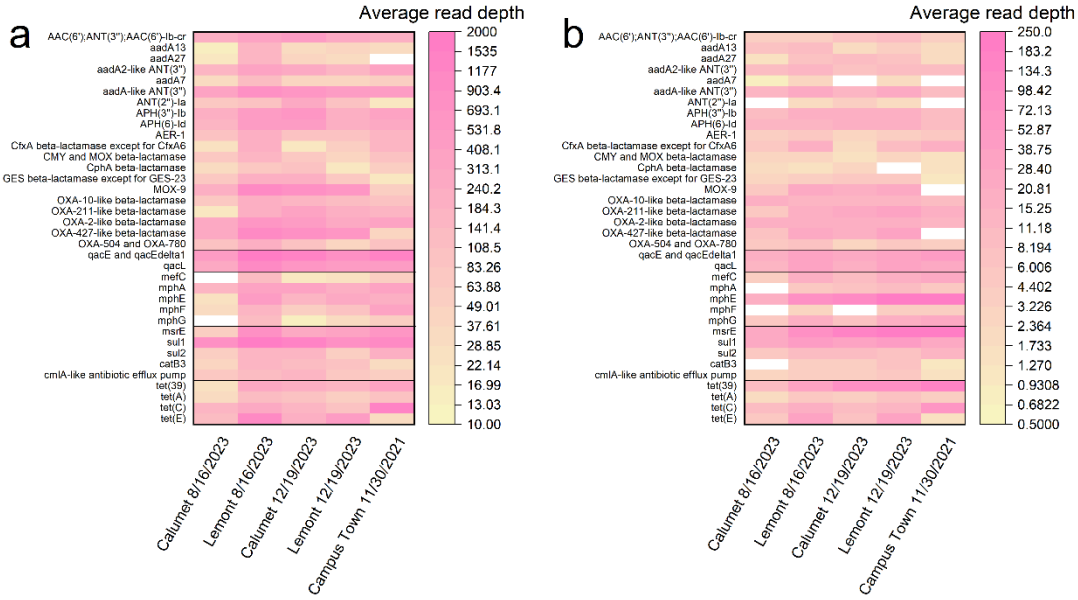

Figure 5. The heatmaps for the most abundant ARGs (i.e., the ARGs with read depths greater than 100 in at least one sample detected by CRISPR-NGS) in five different sewage samples. a: The read depths of the most abundant ARGs detected by CRISPR-NGS libraries; b: The read depths of the most abundant ARGs detected by regular NGS libraries. The read depths increase from yellow to pink. The white color means the ARG was not detected in the sample by the corresponding method. From top to bottom, the heatmaps were divided into seven parts by horizontal black lines according to the drug classes of the corresponding ARGs: aminoglycoside, beta-lactam, antiseptic, macrolide, multi-drug, phenicol, and tetracycline.

Though the profiles of the most abundant ARGs were similar between CRISPR-NGS and regular NGS libraries, when extend the scale to all ARGs with high or low abundances, the profiles detected by CRISPR-NGS and regular NGS were different (**Figure S5**). The entire class of aminocoumarin antibiotics was not detected by regular NGS libraries but detected by CRISPR-NGS libraries. Fewer classes of beta-lactamase genes, for example, TEM, KPC, CMY, CTX-M, and MIR, were detected by regular NGS libraries than CRISPR-NGS libraries. Similarly, regular NGS libraries missed several clusters of multi-drug, peptide, and phenicol resistance genes which were captured by CRISPR-NGS libraries. But for aminoglycoside, macrolide, and tetracycline resistance genes, regular NGS libraries detected similar numbers as CRISPR-NGS libraries. Such findings suggested that regular NGS method is still applicable when aminoglycoside, macrolide, and tetracycline resistance genes are the most interested. But when the range of ARGs for surveillance expands, CRISPR-NGS works better in capturing the ARGs in the classes that are less abundant than the three drug classes mentioned above.

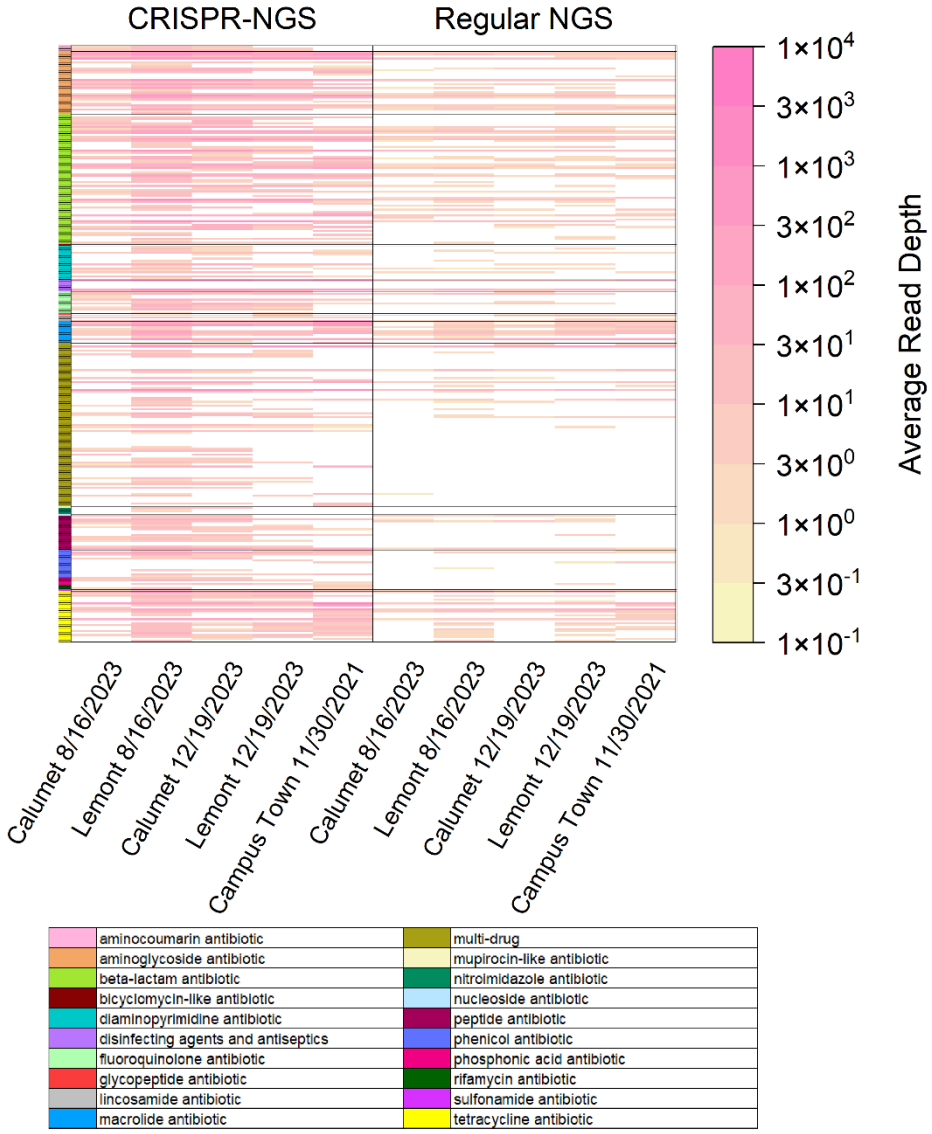

Figure S5. The heatmap for the average read depths of all ARG clusters in all sewage samples measured by CRISPR-NGS method and regular NGS method. The read depths are colored in log-10 scale. The read depths increase from yellow to pink. The ARG clusters are sorted by different drug classes. The drug classes are labeled by different colors on the left.

4 Discussion

Metagenomics has been applied widely in ARG surveillance in different environmental samples. The advantage of metagenomics is that it can detect a wide range of ARGs without knowing the targets (Dai et al., 2022; Davis et al., 2023; Garner et al., 2018; Guo et al., 2017; W. Li et al., 2022; Lira et al., 2020; Majeed et al., 2021; Ng et al., 2017; Talat et al., 2023; Yin et al., 2019). However, it is challenging to sequence ARGs present in low abundances by using metagenomics. We overcome this disadvantage by using CRISPR-Cas9 with multiplexed gRNA to specifically enrich the ARGs in wastewater DNA samples during NGS library preparation.

After the CRISPR-Cas9 treatment, compared to regular NGS libraries, up to 686

more ARGs were detected, and the read depths of the ARGs in the sequencing results increased by 10-fold. Because of the target enrichment step, potential ARGs that have not been recorded in databases may be overlooked in sequencing. Based on the findings in this study, CRISPR-NGS works the best for samples that contain highly diverse ARGs in relatively low abundances. Various types of extended-spectrum beta-lactamase (ESBL) genes were detected in human sewage samples in this study by the CRISPR-NSG method. These genes include not only the common ones like TEM, CTX-M, and OXA, but also the rare ones, such as GES, PER, VEB, BEL, BES, and TLA (Castanheira et al., 2021; Naas et al., 2008). Among these genes, GES, OXA, and VEB were abundant enough to be detected by regular NGS method. Other genes were too diluted to be stably detected without CRISPR treatment before NGS, including the clinically important carbapenem resistance gene KPC. Extra effort is required, and the return is low to monitor these rare ESBLs using conventional methods because their abundances were too low to be detected in most cases. However, studies showed that these ESBLs were also evolving and spreading globally (Gniadkowski, 2008; Pagani et al., 2004). For example, GES-2 gained resistance to carbapenem by only a single point mutation (Poirel et al., 2001). Though rarely detected in the clinical isolates obtained until now, our findings indicated that these ESBL genes were circulating in human communities and could potentially cause public health problems in the future. For example, though considered as a less common ESBL, the read depths of GES in WWTP samples were comparable to OXA, and higher than any other beta-lactamase genes. In contrast, GES was not as abundant as in WWTP samples for the Campus Town manhole sample. This finding suggested that the presence and abundance of different ESBL may be dynamic. Innovative high-throughput detection methods with high sensitivity such as CRISPR-NGS will be helpful in the surveillance of the emergence of such unexpected ARGs.

The CRISPR-NGS method is promising for ARG detection, especially for the surveillance of sewage samples which may have many unknown targets. Early warning of new threats in public health is challenging, because these new targets may be hidden in low abundance in human communities and suddenly emerge to cause an outbreak (Diamond et al., 2022). Surveillance methods should be high in throughput and sensitivity simultaneously. When the throughput is high, multiplexed qPCR is labor-intensive in primer design and sample preparation. Because of such reasons, NGS is more suitable for monitoring unexpected targets than qPCR. An increase in the sensitivity of conventional NGS can be achieved by increasing the total number of reads during a sequencing run (Gweon et al., 2019; Zaheer et al., 2018). However, this method can be costly, because increasing the total read depth by 10-fold means the total sequencing cost for the same sample is increased by 10-fold. Currently, the Illumina sequencing cost is around $2.2 per million reads (Pricing|Biotech). For a library with 50 million reads, the cost solely for sequencing is around $110. If the total number of reads increases by 10-fold to 500 million, the sequencing cost for the same library will increase to around $1,100. Considering more than 99.9% reads are irrelevant to ARG detection (Guitor et al., 2020), increasing the proportion of ARG reads before loading the samples to sequencing machine will be more economical. CRISPR-NGS method could increase the ARG read depth by around 10-fold with the same total number of reads in sequencing with only an increased cost of around $30 in Cas9 nuclease and guide RNA preparation for each library. The relatively low cost makes the CRISPR-NGS method attractive for routine ARG surveillance in the future.

There are also limitations of the CRISPR-NGS method. First, the library preparation time for CRISPR-NGS method is longer than regular NGS method. Conventional NGS library preparation methods have been optimized and commercialized: libraries can be prepared in a few hours. More steps and longer incubation time are required in CRISPR-NGS library preparation, including guide RNA preparation and Cas9 cleavage. It usually

requires two days to complete the entire protocol of CRISPR-NGS library preparation. Second, genetic information other than ARGs, for example, mutations on the genes encoding the proteins that are usually targeted by drugs, may be missed by CRISPR-NGS method, because these sequences are not targeted during assay design. ARGs that are currently unknown may also be missed for the same reason. Third, CRISPR-NGS is not as accurate for ARG quantification as qPCR. We recommend using CRISPR-NGS periodically for environmental surveillance to identify ARGs relevant for public health and use qPCR to quantify these ARGs.

5 Conclusions

Here, we introduced a CRISPR-Cas9 modified NGS library preparation method to increase the sensitivity in ARG detection in sewage samples. Up to 29 ARG families or up to 686 ARGs present in low abundance were uncovered by NGS after CRISPR-Cas9 modification. An average of 10-fold increases in read depths for ARGs were achieved. With the same total read depth for one sequencing run, CRISPR-NGS libraries can achieve a better sensitivity in ARG detection with comparable throughput as regular NGS.

Acknowledgement

We acknowledge funding from Water Research Foundation and USEPA. We acknowledge the support from Dr. Xing Wang, Lucas D. Akin, and Abhisek Dwivedy. We acknowledge the effort in sample collection from Arthur R. Schmidt IV, Ryan McAllister, Amaja Craft, and Josie Hoppenworth. We acknowledge Dr. Rachel J. Whitaker, Dr. Jayadevi H. Chandrashekhar, and Laura C. Suttenfield in sharing *Pseudomonas aeruginosa* PA14. We acknowledge Roy J. Carver Biotechnology Center in their sequencing services. We acknowledge Ashley Ceniceros, Ann Frederick, and Bruce Wellman from Carle Health for providing clinical bacteria isolates. We acknowledge Dr. Albert Cox, Essam El-Naggar, Kamlesh Patel, and Kaylyn Patterson from Metropolitan Water Reclamation District of Greater Chicago for providing WWTP sewage samples. We acknowledge Dr. Awais Vaid and Kip Stevenson for the guidance in manhole sewage sample collection.

References

Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., Edalatmand, A., Petkau, A., Syed, S. A., Tsang, K. K., Baker, S. J. C., Dave, M., Mccarthy, M. C., Mukiri, K. M., Nasir, J. A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., … Mcarthur, A. G. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research*, *51*(D1), D690–D699. https://doi.org/10.1093/NAR/GKAC920

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A. L. V., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H. K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., … McArthur, A. G. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, *48*(D1), D517–D525. https://doi.org/10.1093/NAR/GKZ935

Anjum, M. F., Zankari, E., & Hasman, H. (2017). Molecular Methods for Detection of Antimicrobial Resistance. *Microbiology Spectrum*, *5*(6). https://doi.org/10.1128/MICROBIOLSPEC.ARBA-0011-2017

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., … Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, *9*(1), 75. https://doi.org/10.1186/1471-2164-9-75

Bengtsson-Palme, J., Abramova, A., Berendonk, T. U., Coelho, L. P., Forslund, S. K., Gschwind, R., Heikinheimo, A., Jarquín-Díaz, V. H., Khan, A. A., Klümper, U., Löber, U., Nekoro, M., Osińska, A. D., Ugarcina Perovic, S., Pitkänen, T., Rødland, E. K., Ruppé, E., Wasteson, Y., Wester, A. L., & Zahra, R. (2023). Towards monitoring of antimicrobial resistance in the environment: For what reasons, how to implement it, and what are the data needs? *Environment International*, *178*, 108089. https://doi.org/10.1016/J.ENVINT.2023.108089

Bengtsson-Palme, J., Kristiansson, E., & Larsson, D. G. J. (2018). Environmental factors influencing the development and spread of antibiotic resistance. *FEMS Microbiology Reviews*, *42*(1), 68–80. https://doi.org/10.1093/FEMSRE/FUX053

Burnham, C. A. D., Leeds, J., Nordmann, P., O'Grady, J., & Patel, J. (2017). Diagnosing antimicrobial resistance. *Nature Reviews Microbiology 2017 15:11*, *15*(11), 697–703. https://doi.org/10.1038/nrmicro.2017.103

Castanheira, M., Simner, P. J., & Bradford, P. A. (2021). Extended-spectrum β-lactamases: an update on their characteristics, epidemiology and detection. *JAC-Antimicrobial Resistance*, *3*(3). https://doi.org/10.1093/JACAMR/DLAB092

Dai, D., Brown, C., Bürgmann, H., Larsson, D. G. J., Nambi, I., Zhang, T., Flach, C. F., Pruden, A., & Vikesland, P. J. (2022). Long-read metagenomic sequencing reveals shifts in associations of antibiotic resistance genes with mobile genetic elements from sewage to activated sludge. *Microbiome*, *10*(1), 1–16. https://doi.org/10.1186/S40168-021-01216-5/FIGURES/6

Davis, B. C., Brown, C., Gupta, S., Calarco, J., Liguori, K., Milligan, E., Harwood, V. J., Pruden, A., & Keenum, I. (2023). Recommendations for the use of metagenomics for routine monitoring of antibiotic resistance in wastewater and impacted aquatic environments. *Critical Reviews in Environmental Science and Technology*, *53*(19), 1731–1756. https://doi.org/10.1080/10643389.2023.2181620

de Abreu, V. A. C., Perdigão, J., & Almeida, S. (2020). Metagenomic Approaches to Analyze Antimicrobial Resistance: An Overview. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/FGENE.2020.575592

Debroas, D., & Siguret, C. (2019). Viruses as key reservoirs of antibiotic resistance genes in the environment. *The ISME Journal*, *13*(11), 2856. https://doi.org/10.1038/S41396-019-0478-9

Diamond, M. B., Keshaviah, A., Bento, A. I., Conroy-Ben, O., Driver, E. M., Ensor, K. B., Halden, R. U., Hopkins, L. P., Kuhn, K. G., Moe, C. L., Rouchka, E. C., Smith, T., Stevenson, B. S., Susswein, Z., Vogel, J. R., Wolfe, M. K., Stadler, L. B., & Scarpino, S. V. (2022). Wastewater surveillance of pathogens can inform public health responses. *Nature Medicine 2022 28:10*, *28*(10), 1992–1995. https://doi.org/10.1038/s41591-022-01940-x

Duncavage, E. J., Magrini, V., Becker, N., Armstrong, J. R., Demeter, R. T., Wylie, T., Abel, H. J., & Pfeifer, J. D. (2011). Hybrid Capture and Next-Generation Sequencing Identify Viral

Integration Sites from Formalin-Fixed, Paraffin-Embedded Tissue. *The Journal of Molecular Diagnostics*, *13*(3), 325–333. https://doi.org/10.1016/J.JMOLDX.2011.01.006

Flint, A., Cooper, A., Rao, M., Weedmark, K., Carrillo, C., & Tamber, S. (2023). Targeted metagenomics using bait-capture to detect antibiotic resistance genes in retail meat and seafood. *Frontiers in Microbiology*, *14*, 1188872. https://doi.org/10.3389/FMICB.2023.1188872/BIBTEX

Garner, E., Chen, C., Xia, K., Bowers, J., Engelthaler, D. M., McLain, J., Edwards, M. A., & Pruden, A. (2018). Metagenomic Characterization of Antibiotic Resistance Genes in Full-Scale Reclaimed Water Distribution Systems and Corresponding Potable Systems. *Environmental Science and Technology*, *52*(11), 6113–6125. https://doi.org/10.1021/ACS.EST.7B05419/SUPPL_FILE/ES7B05419_SI_004.XLSX

Gilpatrick, T., Lee, I., Graham, J. E., Raimondeau, E., Bowen, R., Heron, A., Downs, B., Sukumar, S., Sedlazeck, F. J., & Timp, W. (2020). Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nature Biotechnology 2020 38:4*, *38*(4), 433–438. https://doi.org/10.1038/s41587-020-0407-5

Ginzinger, D. G. (2002). Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. *Experimental Hematology*, *30*(6), 503–512. https://doi.org/10.1016/S0301-472X(02)00806-8

Gniadkowski, M. (2008). Evolution of extended-spectrum beta-lactamases by mutation. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *14 Suppl 1*(SUPPL. 1), 11–32. https://doi.org/10.1111/J.1469-0691.2007.01854.X

Grace, A., Sahu, R., Owen, D. R., & Dennis, V. A. (2022). Pseudomonas aeruginosa reference strains PAO1 and PA14: A genomic, phenotypic, and therapeutic review. *Frontiers in Microbiology*, *13*, 1023523. https://doi.org/10.3389/FMICB.2022.1023523/BIBTEX

Guitor, A. K., Raphenya, A. R., Klunk, J., Kuch, M., Alcock, B., Surette, M. G., McArthur, A. G., Poinar, H. N., & Wright, G. D. (2020). Capturing the Resistome: a Targeted Capture Method To Reveal Antibiotic Resistance Determinants in Metagenomes. *Antimicrobial Agents and Chemotherapy*, *64*(1). https://doi.org/10.1128/AAC.01324-19

Guo, J., Li, J., Chen, H., Bond, P. L., & Yuan, Z. (2017). Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. *Water Research*, *123*, 468–478. https://doi.org/10.1016/J.WATRES.2017.07.002

Gweon, H. S., Shaw, L. P., Swann, J., De Maio, N., Abuoun, M., Niehus, R., Hubbard, A. T. M., Bowes, M. J., Bailey, M. J., Peto, T. E. A., Hoosdally, S. J., Walker, A. S., Sebra, R. P., Crook, D. W., Anjum, M. F., Read, D. S., Stoesser, N., Abuoun, M., Anjum, M., … Woodford, N. (2019). The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. *Environmental Microbiomes*, *14*(1), 1–15. https://doi.org/10.1186/S40793-019-0347-1/FIGURES/7

Han, R., Shi, Q., Wu, S., Yin, D., Peng, M., Dong, D., Zheng, Y., Guo, Y., Zhang, R., & Hu, F. (2020). Dissemination of Carbapenemases (KPC, NDM, OXA-48, IMP, and VIM) Among Carbapenem-Resistant Enterobacteriaceae Isolated From Adult and Children Patients in China. *Frontiers in Cellular and Infection Microbiology*, *10*, 535376. https://doi.org/10.3389/FCIMB.2020.00314/BIBTEX

Hayden, H. S., Joshi, S., Radey, M. C., Vo, A. T., Forsberg, C., Morgan, S. J., Waalkes, A., Holmes, E. A., Klee, S. M., Emond, M. J., Singh, P. K., & Salipante, S. J. (2022). Genome Capture Sequencing Selectively Enriches Bacterial DNA and Enables Genome-Wide Measurement of Intrastrain Genetic Diversity in Human Infections. *MBio*, *13*(5). https://doi.org/10.1128/MBIO.01424-22/SUPPL_FILE/MBIO.01424-22-S0009.DOCX

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications 2018 9:1*, *9*(1), 1–8. https://doi.org/10.1038/s41467-018-07641-9

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology 2019 37:5*, *37*(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/GR.215087.116/-/DC1

Kowalska-Krochmal, B., & Dudek-Wicher, R. (2021). The Minimum Inhibitory Concentration of Antibiotics: Methods, Interpretation, Clinical Relevance. *Pathogens*, *10*(2), 1–21. https://doi.org/10.3390/PATHOGENS10020165

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods 2012 9:4*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Larsson, D. G. J., & Flach, C. F. (2021). Antibiotic resistance in the environment. *Nature Reviews Microbiology 2021 20:5*, *20*(5), 257–269. https://doi.org/10.1038/s41579-021-00649-x

Le, T. H., Ng, C., Chen, H., Yi, X. Z., Koh, T. H., Barkham, T. M. S., Zhou, Z., & Gin, K. Y. H. (2016). Occurrences and Characterization of Antibiotic-Resistant Bacteria and Genetic Determinants of Hospital Wastewater in a Tropical Country. *Antimicrobial Agents and Chemotherapy*, *60*(12), 7449. https://doi.org/10.1128/AAC.01556-16

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/BIOINFORMATICS/BTP352

Li, W., Mao, F., Ng, C., Jong, M. C., Goh, S. G., Charles, F. R., Ng, O. T., Marimuthu, K., He, Y., & Gin, K. Y. H. (2022). Population-based variations of a core resistome revealed by urban sewage metagenome surveillance. *Environment International*, *163*, 107185. https://doi.org/10.1016/J.ENVINT.2022.107185

Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S., Ravinder, N., & Chesnut, J. D. (2015). Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of Biotechnology*, *208*, 44–53. https://doi.org/10.1016/J.JBIOTEC.2015.04.024

Liguori, K., Keenum, I., Davis, B. C., Calarco, J., Milligan, E., Harwood, V. J., & Pruden, A. (2022). Antimicrobial Resistance Monitoring of Water Environments: A Framework for Standardized Methods and Quality Control. *Environmental Science and Technology*, *56*(13), 9149–9160. https://doi.org/10.1021/ACS.EST.1C08918/ASSET/IMAGES/LARGE/ES1C08918_0005

.JPEG

Lira, F., Vaz-Moreira, I., Tamames, J., Manaia, C. M., & Martínez, J. L. (2020). Metagenomic analysis of an urban resistome before and after wastewater treatment. *Scientific Reports 2020 10:1*, *10*(1), 1–9. https://doi.org/10.1038/s41598-020-65031-y

Lund, M., Petersen, M. B., Jørgensen, A. L., Paulmann, D., & Wang, M. (2018). Rapid real-time PCR for the detection of IMP, NDM, VIM, KPC and OXA-48 carbapenemase genes in isolates and spiked stool samples. *Diagnostic Microbiology and Infectious Disease*, *92*(1), 8–12. https://doi.org/10.1016/J.DIAGMICROBIO.2018.04.002

Maboni, G., Baptista, R. de P., Wireman, J., Framst, I., Summers, A. O., & Sanchez, S. (2022). Three Distinct Annotation Platforms Differ in Detection of Antimicrobial Resistance Genes in Long-Read, Short-Read, and Hybrid Sequences Derived from Total Genomic DNA or from Purified Plasmid DNA. *Antibiotics*, *11*(10), 1400. https://doi.org/10.3390/ANTIBIOTICS11101400/S1

Majeed, H. J., Riquelme, M. V., Davis, B. C., Gupta, S., Angeles, L., Aga, D. S., Garner, E., Pruden, A., & Vikesland, P. J. (2021). Evaluation of Metagenomic-Enabled Antibiotic Resistance Surveillance at a Conventional Wastewater Treatment Plant. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/FMICB.2021.657954

McCarty, N. S., Graham, A. E., Studená, L., & Ledesma-Amaro, R. (2020). Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nature Communications 2020 11:1*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-15053-x

McCombie, W. R., & McPherson, J. D. (2019). Future Promises and Concerns of Ubiquitous Next-Generation Sequencing. *Cold Spring Harbor Perspectives in Medicine*, *9*(9). https://doi.org/10.1101/CSHPERSPECT.A025783

McGinnis, S., & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, *32*(Web Server issue), W20. https://doi.org/10.1093/NAR/GKH435

McKenna, M. (2013). Antibiotic resistance: The last resort. *Nature*, *499*(7459), 394–396. https://doi.org/10.1038/499394A

McLain, J. E., Cytryn, E., Durso, L. M., & Young, S. (2016). Culture-based Methods for Detection of Antibiotic Resistance in Agroecosystems: Advantages, Challenges, and Gaps in Knowledge. *Journal of Environmental Quality*, *45*(2), 432–440. https://doi.org/10.2134/JEQ2015.06.0317

Muyzer, G., De Waal, E. C., & Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*, *59*(3), 695–700. https://doi.org/10.1128/AEM.59.3.695-700.1993

Naas, T., Poirel, L., & Nordmann, P. (2008). Minor extended-spectrum β-lactamases. *Clinical Microbiology and Infection*, *14*(SUPPL. 1), 42–52. https://doi.org/10.1111/J.1469-0691.2007.01861.X

Ng, C., Tay, M., Tan, B., Le, T. H., Haller, L., Chen, H., Koh, T. H., Barkham, T. M. S., & Gin, K. Y. H. (2017). Characterization of metagenomes in urban aquatic compartments reveals high prevalence of clinically relevant antibiotic resistance genes in wastewaters. *Frontiers in Microbiology*, *8*(NOV), 298023. https://doi.org/10.3389/FMICB.2017.02200/BIBTEX

Olson, R. D., Assaf, R., Brettin, T., Conrad, N., Cucinell, C., Davis, J. J., Dempsey, D. M., Dickerman, A., Dietrich, E. M., Kenyon, R. W., Kuscuoglu, M., Lefkowitz, E. J., Lu, J., Machi, D., Macken, C., Mao, C., Niewiadomska, A., Nguyen, M., Olsen, G. J., … Stevens, R. L. (2023). Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Research*, *51*(D1), D678–D689. https://doi.org/10.1093/NAR/GKAC1003

Pagani, L., Mantengoli, E., Migliavacca, R., Nucleo, E., Pollini, S., Spalla, M., Daturi, R., Romero, E., & Rossolini, G. M. (2004). Multifocal Detection of Multidrug-Resistant Pseudomonas aeruginosa Producing the PER-1 Extended-Spectrum β-Lactamase in Northern Italy. *Journal of Clinical Microbiology*, *42*(6), 2523. https://doi.org/10.1128/JCM.42.6.2523-2529.2004

Poirel, L., Weldhagen, G. F., Naas, T., De Champs, C., Dove, M. G., & Nordmann, P. (2001). GES-2, a class A beta-lactamase from Pseudomonas aeruginosa with increased hydrolysis of imipenem. *Antimicrobial Agents and Chemotherapy*, *45*(9), 2598–2603. https://doi.org/10.1128/AAC.45.9.2598-2603.2001

Quan, J., Langelier, C., Kuchta, A., Batson, J., Teyssier, N., Lyden, A., Caldera, S., McGeever, A., Dimitrov, B., King, R., Wilheim, J., Murphy, M., Ares, L. P., Travisano, K. A., Sit, R., Amato, R., Mumbengegwi, D. R., Smith, J. L., Bennett, A., … Crawford, E. D. (2019). FLASH: a next-generation CRISPR diagnostic for multiplexed detection of antimicrobial resistance sequences. *Nucleic Acids Research*, *47*(14), E83. https://doi.org/10.1093/NAR/GKZ418

Rao, X., Huang, X., Zhou, Z., & Lin, X. (2013). An improvement of the 2ˆ(–delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. *Biostatistics, Bioinformatics and Biomathematics*, *3*(3), 71. /pmc/articles/PMC4280562/

Razzaque, M. S. (2021). Implementation of antimicrobial stewardship to reduce antimicrobial drug resistance. *Expert Review of Anti-Infective Therapy*, *19*(5), 559–562. https://doi.org/10.1080/14787210.2021.1840977

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative Genomics Viewer. *Nature Biotechnology*, *29*(1), 24. https://doi.org/10.1038/NBT.1754

Rochegüe, T., Haenni, M., Cazeau, G., Metayer, V., Madec, J. Y., Ferry, T., & Lupo, A. (2021). An inventory of 44 qPCR assays using hydrolysis probes operating with a unique amplification condition for the detection and quantification of antibiotic resistance genes. *Diagnostic Microbiology and Infectious Disease*, *100*(3), 115328. https://doi.org/10.1016/J.DIAGMICROBIO.2021.115328

Roschanski, N., Fischer, J., Guerra, B., & Roesler, U. (2014). Development of a Multiplex Real-Time PCR for the Rapid Detection of the Predominant Beta-Lactamase Genes CTX-M, SHV, TEM and CIT-Type AmpCs in Enterobacteriaceae. *PLOS ONE*, *9*(7), e100956. https://doi.org/10.1371/JOURNAL.PONE.0100956

Ruby, J. G., Bellare, P., & DeRisi, J. L. (2013). PRICE: Software for the targeted assembly of components of (Meta) genomic sequence data. *G3: Genes, Genomes, Genetics*, *3*(5), 865–880. https://doi.org/10.1534/G3.113.005967/-/DC1/FILES1.FASTA

Samorodnitsky, E., Jewell, B. M., Hagopian, R., Miya, J., Wing, M. R., Lyon, E., Damodaran, S., Bhatt, D., Reeser, J. W., Datta, J., & Roychowdhury, S. (2015). Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing.

Human Mutation, *36*(9), 903–914. https://doi.org/10.1002/HUMU.22825

Samreen, Ahmad, I., Malak, H. A., & Abulreesh, H. H. (2021). Environmental antimicrobial resistance and its drivers: a potential threat to public health. *Journal of Global Antimicrobial Resistance*, *27*, 101–111. https://doi.org/10.1016/J.JGAR.2021.08.001

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11), 2498. https://doi.org/10.1101/GR.1239303

Shen, C.-H. (2023). Amplification of nucleic acids. *Diagnostic Molecular Biology*, 259–293. https://doi.org/10.1016/B978-0-323-91788-9.00010-7

Skandalis, N., Maeusli, M., Papafotis, D., Miller, S., Lee, B., Theologidis, I., & Luna, B. (2021). Environmental Spread of Antibiotic Resistance. *Antibiotics (Basel, Switzerland)*, *10*(6). https://doi.org/10.3390/ANTIBIOTICS10060640

Smith, T., Heger, A., & Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Research*, *27*(3), 491–499. https://doi.org/10.1101/GR.209601.116

Srathongneam, T., Sresung, M., Paisantham, P., Ruksakul, P., Singer, A. C., Sukchawalit, R., Satayavivad, J., Mongkolsuk, S., & Sirikanchana, K. (2024). High throughput qPCR unveils shared antibiotic resistance genes in tropical wastewater and river water. *Science of The Total Environment*, *908*, 167867. https://doi.org/10.1016/J.SCITOTENV.2023.167867

Sundararaman, B., Sylvester, M. D., Kozyreva, V. K., Berrada, Z. L., Corbett-Detig, R. B., & Green, R. E. (2023). A hybridization target enrichment approach for pathogen genomics. *MBio*, *14*(5). https://doi.org/10.1128/MBIO.01889-23

Talat, A., Blake, K. S., Dantas, G., & Khan, A. U. (2023). Metagenomic Insight into Microbiome and Antibiotic Resistance Genes of High Clinical Concern in Urban and Rural Hospital Wastewater of Northern India Origin: a Major Reservoir of Antimicrobial Resistance. *Microbiology Spectrum*, *11*(2). https://doi.org/10.1128/SPECTRUM.04102-22/ASSET/EBEB3182-2F24-4F24-9DF0-869DFE06D732/ASSETS/IMAGES/MEDIUM/SPECTRUM.04102-22-F009.GIF

United Nations Environment Programme. (2023). *Bracing for superbugs : strengthening environmental action in the One Health response to antimicrobial resistance.*

Vaser, R., & Šikić, M. (2021). Time- and memory-efficient genome assembly with Raven. *Nature Computational Science 2021 1:5*, *1*(5), 332–336. https://doi.org/10.1038/s43588-021-00073-4

Wick, R. R., Judd, L. M., Cerdeira, L. T., Hawkey, J., Méric, G., Vezina, B., Wyres, K. L., & Holt, K. E. (2021). Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biology*, *22*(1), 1–17. https://doi.org/10.1186/S13059-021-02483-Z/FIGURES/4

Xie, N. G., Wang, M. X., Song, P., Mao, S., Wang, Y., Yang, Y., Luo, J., Ren, S., & Zhang, D. Y. (2022). Designing highly multiplex PCR primer sets with Simulated Annealing Design using Dimer Likelihood Estimation (SADDLE). *Nature Communications 2022 13:1*, *13*(1), 1–10. https://doi.org/10.1038/s41467-022-29500-4

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-

BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*, 134. https://doi.org/10.1186/1471-2105-13-134

Yin, X., Deng, Y., Ma, L., Wang, Y., Chan, L. Y. L., & Zhang, T. (2019). Exploration of the antibiotic resistome in a wastewater treatment plant by a nine-year longitudinal metagenomic study. *Environment International*, *133*, 105270. https://doi.org/10.1016/J.ENVINT.2019.105270

Zaheer, R., Noyes, N., Ortega Polo, R., Cook, S. R., Marinier, E., Van Domselaar, G., Belk, K. E., Morley, P. S., & McAllister, T. A. (2018). Impact of sequencing depth on the characterization of the microbiome and resistome. *Scientific Reports 2018 8:1*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-24280-8

Zhang, Q., & Ishii, S. (2018). Improved simultaneous quantification of multiple waterborne pathogens and fecal indicator bacteria with the use of a sample process control. *Water Research*, *137*, 193–200. https://doi.org/10.1016/J.WATRES.2018.03.023